UNIVERSITY OF CAMBRIDGE
Faculty of Economics

Institute for New Economic Thinking

Cambridge Working Papers in Economics

Cambridge-INET Working Paper Series No: 2019/10

Cambridge Working Papers in Economics: 1952

DEPENDENT MICROSTRUCTURE NOISE AND INTEGRATED VOLATILITY ESTIMATION FROM HIGH-FREQUENCY DATA

Z. Merrick Li (University of Cambridge)

Roger J. A. Laeven (University of Amsterdam)

Michel H. Vellekoop (University of Amsterdam)

In this paper, we develop econometric tools to analyze the integrated volatility (IV) of the efficient price and the dynamic properties of microstructure noise in high-frequency data under general dependent noise. We first develop consistent estimators of the variance and autocovariances of noise using a variant of realized volatility. Next, we employ these estimators to adapt the pre-averaging method and derive consistent estimators of the IV, which converge stably to a mixed Gaussian distribution at the optimal rate $n^{1/4}$. To improve the _nite sample performance, we propose a multi-step approach that corrects the finite sample bias, which turns out to be crucial in applications. Our extensive simulation studies demonstrate the excellent performance of our multi-step estimators. In an empirical study, we analyze the dependence structures of microstructure noise and provide intuitive economic interpretations; we also illustrate the importance of accounting for both the serial dependence in noise and the finite sample bias when estimating IV.

Dependent Microstructure Noise and Integrated Volatility Estimation from High-Frequency Data

Z. Merrick Li*

Department of Economics

University of Cambridge

Roger J. A. Laeven†

Amsterdam School of Economics

University of Amsterdam, EURANDOM and CentER

Michel H. Vellekoop‡

Amsterdam School of Economics

University of Amsterdam

June 8, 2019

Abstract

In this paper, we develop econometric tools to analyze the integrated volatility (IV) of the efficient price and the dynamic properties of microstructure noise in high-frequency data under general dependent noise. We first develop consistent estimators of the variance and autocovariances of noise using a variant of realized volatility. Next, we employ these estimators to adapt the pre-averaging method and derive consistent estimators of the IV, which converge stably to a mixed Gaussian distribution at the optimal rate $n^{1/4}$. To improve the finite sample performance, we propose a multi-step approach that corrects the finite sample bias, which turns out to be crucial in applications. Our extensive simulation studies demonstrate the excellent performance of our multi-step estimators. In an empirical study, we analyze the dependence structures of microstructure noise and provide intuitive economic interpretations; we also illustrate the importance of accounting for both the serial dependence in noise and the finite sample bias when estimating IV.

Keywords: Dependent microstructure noise, realized volatility, bias correction, integrated volatility, mixing sequences, pre-averaging method.

JEL classification: C13, C14, C55, C58.

*Corresponding author. Department of Economics, University of Cambridge, Austin Robinson Building, Sidgwick Avenue, Cambridge, CB3 9DD, United Kingdom. Email: Z.Merrick.Li@gmail.com. Phone: +44 (0) 1223 335288.

†University of Amsterdam, Amsterdam School of Economics, PO Box 15867, 1001 NJ Amsterdam, The Netherlands. Email: R.J.A.Laeven@uva.nl. Phone: +31 (0)20 5254219.

‡University of Amsterdam, Amsterdam School of Economics, PO Box 15867, 1001 NJ Amsterdam, The Netherlands. Email: M.H.Vellekoop@uva.nl. Phone: +31 (0)20 5254210.

1 Introduction

Over the past decade and a half, high-frequency financial data have become increasingly available. In tandem, the development of econometric tools to study the dynamic properties of high-frequency data has become an important subject area in economics and statistics. A major challenge is provided by the accumulation of market microstructure noise at higher frequencies, which can be attributed to various market microstructure effects including, for example, information asymmetries (see Glosten and Milgrom (1985)), inventory controls (see Ho and Stoll (1981)), discreteness of the data (see Harris (1990)), and transaction costs (see Garman (1976)).

It has been well-established (see, e.g., Black (1986)) that the observed transaction price[1] Y can be decomposed into the unobservable "efficient price" (or "frictionless equilibrium price") X plus a noise component U that captures market microstructure effects. That is, it is natural to assume that

$$Y_t = X_t + U_t, \tag{1}$$

where further assumptions on X and U need to be stipulated. While estimating the IV of the efficient price is a canonical problem in high-frequency financial econometrics (see, for example, Aït-Sahalia and Jacod (2014)), the study of microstructure noise, e.g., its magnitude, dynamic properties, etc., is the main focus of the market microstructure literature (see, for example, Hasbrouck (2007)). A common challenge, however, is that the two components of the observed price Y in (1) are latent. Therefore, distributional features of one component, say, of the microstructure noise, will affect the estimation of characteristics of the other, such as the IV of the efficient price.[2]

While the semimartingale framework provides the natural class to model the efficient price (see, e.g., Duffie (2010)), the statistical assumptions on noise induced by microeconomic financial models range from simple to very complex, depending on which phenomena the model aims to capture. For example, the classic Roll model (see Roll (1984)) postulates an i.i.d. bid-ask bounce resulting from uncorrelated order flows; Hasbrouck and Ho (1987), Choi et al. (1988), and Stoll (1989) introduce autocorrelated order flows, yielding autoregressive microstructure noise; and Gross-KlussMann and Hautsch (2013) model microstructure noise with long-memory properties. Therefore, being able to account for the potentially complex statistical behavior of microstructure noise that contaminates our observations of the semimartingale efficient price dynamics, would be an appealing property of any method that aims at disentangling the efficient price and microstructure noise.

[1] In this paper, "price" always refers to the "logarithmic price".

[2] Indeed, while high-frequency data in principle facilitate the asymptotic and empirical analysis of volatility estimators, the pronounced presence of microstructure noise at high frequency subverts the desirable properties of traditional estimators such as realized volatility.

To estimate the IV of the efficient price, several de-noise methods have been developed, mostly assuming i.i.d. microstructure noise. Examples include the two-scale and multi-scale realized volatility estimators developed in Zhang et al. (2005) and Zhang (2006), the likelihood approach initiated by Aït-Sahalia et al. (2005) and Xiu (2010), the realized kernel methods developed in Barndorff-Nielsen et al. (2008), and the pre-averaging method developed in a series of papers by Podolskij and Vetter (2009b) and Jacod et al. (2009, 2010), see also Podolskij and Vetter (2009a). The variance of noise is usually obtained as a by-product.

In this paper, we allow the microstructure noise to be serially dependent in a general setting, nesting many special cases (including independence). We do not impose any parametric restrictions on the distribution of the noise, except for some rather general mixing conditions that guarantee the existence of limit distributions, hence our approach is essentially nonparametric. In this setting, we first derive the stochastic limit of the realized volatility of observed prices after j lags. Using this limit result, we develop consistent estimators of the variance and covariances of noise. The aim of estimating the second moments of noise is twofold. On the one hand, we would like to explore the dynamic properties of microstructure noise. In particular, we would like to compare these properties to those induced by various parametric models of microstructure noise based on leading microstructure theory, and obtain corresponding economic interpretations to achieve a better understanding of the microstructure effects in high-frequency data. On the other hand, the second moments of noise become nuisance parameters when estimating the IV, which is a prime objective in the analysis of high-frequency financial data.

To estimate the IV, we next adapt the pre-averaging approach (PAV) to allow for serially dependent noise in our general setting, first based on non-overlapping sampling blocks and next based on overlapping sampling blocks, in both cases using general weight functions (i.e., general kernels). We find that the stochastic limits of the adapted PAV estimators are functions of the volatility and the variance and covariances of noise, and the latter, constituting an *asymptotic bias*, can be consistently estimated by our realized volatility estimator. Hence, we can correct the asymptotic bias, resulting in centered estimators of the IV, for which we establish the associated central limit theorems.

A key interest in this paper is to unravel the interplay between asymptotic and finite sample biases when estimating the IV. In a formal finite sample analysis, we find that the realized volatility estimator has a finite sample bias that is proportional to the IV. This bias term becomes significant when the number of lags (in computing the variant of realized volatility) is large, or the noise-to-signal ratio[3] is small. Therefore, we are in a situation in which the IV generates a *finite sample bias* to the estimators of the second moments of noise, while the latter introduce an *asymptotic bias* when estimating the former. This "feedback effect" in the bias corrections motivates us to develop *multi-step estimators*.

[3]That is, the ratio of the variance of noise and the IV.

First, we simply ignore the dependence in noise and proceed with the pre-averaging method to obtain an estimator of the IV. Next, we use this estimator to obtain *finite sample bias* corrected estimators of the second moments of noise, which can then be used to correct the asymptotic bias, yielding the second-step estimator of the IV. Repeating this process leads to three-step estimators (and beyond). Figure 1 gives a simple graphical illustration of the implementation of the multi-step estimators. We establish consistency and a central limit theorem for our multi-step estimators.

We conduct extensive Monte Carlo experiments to examine the performance of our estimators, which proves to be excellent. We demonstrate in particular that they can accommodate both serially dependent and independent noise and perform well in finite samples with realistic data frequencies and sample sizes. The experiments reveal the importance of a unified treatment of asymptotic and finite sample biases when estimating IV.

Empirically, we apply our new estimators to a sample of Citigroup transaction data. We find that the associated microstructure noise tends to be positively autocorrelated. This is in line with earlier findings in the microstructure literature, see Hasbrouck and Ho (1987), Choi et al. (1988), and Huang and Stoll (1997). When we attribute this positive autocorrelation to order flow continuation, the estimated probability that a buy (or sell) order follows another buy (or sell) order is found to be 0.87. Furthermore, microstructure noise turns out to be negatively autocorrelated under tick time sampling. This is consistent with inventory models, in which dealers alternate quotes to maintain their inventory position. We obtain an estimate of the probability of reversed orders equal to 0.84. Turning to the estimators of IV, we find that with positively autocorrelated noise the commonly adopted methods that hinge on the i.i.d. assumption of noise tend to overestimate the IV. Under two alternative (sub)sampling schemes our estimators also appear to work well. This testifies to the critical relevance of the bias corrections embedded in our multi-step estimators.

In earlier literature, Aït-Sahalia et al. (2011) show that the two-scale and multi-scale realized volatility estimators are robust to exponentially decaying dependence in noise. In this paper, we provide explicit estimators of the second moments of noise and analyze their asymptotic behavior, develop bias-corrected estimators of the IV based on these moments of noise, and empirically assess the noise characteristics. Furthermore, Hautsch and Podolskij (2013) study q-dependent microstructure noise, develop consistent estimators of the first q autocovariances of microstructure noise and define the associated pre-averaging estimators. An appealing feature of their approach is that their autocovariance-type estimators of q-dependent noise consider non-overlapping increments which avoids finite sample bias. We allow for more general assumptions on the dependence structure of microstructure noise. Owing to its generality our setting incorporates many microstructure models as special cases. We therefore do not need to advocate any particular model of microstructure noise.

In two contemporaneous works, Jacod et al. (2017, 2019) also study dependent noise in high-frequency data. In Jacod et al. (2017), they develop a novel local averaging method to "recover" the noise and can, in principle, estimate any finite (joint) moments of noise with diurnal features. Moreover, they also allow observation times to be random. Empirically, they find some interesting statistical properties of noise. In particular, they find that noise is strongly serially dependent, with polynomially decaying autocorrelations. Employing this local averaging method, Jacod et al. (2019) develop an estimator of IV that allows for dependent noise. The local averaging method differs from, and allows to analyze more general noise characteristics than, the simpler realized volatility method developed here. The key difference is our explicit treatment of the feedback effect between the asymptotic and finite sample biases: we show that in a finite sample, the IV and second moments of microstructure noise should be estimated in a unified way, since they induce biases in each other. We design novel and easily implementable multi-step estimators to correct for the intricate biases. Our multi-step estimators of the IV, which are designed to allow for dependent noise, also perform well in the special case of independent noise, and in a sample of reasonable size as encountered in practice. This robustness to (mis)specification of noise and to sampling frequencies is an important advantage of our multi-step estimators. Our unified treatment of the asymptotic and finite sample biases may help explain why the empirical studies in Jacod et al. (2017) render the strong dependence in noise they find (and question themselves); see our empirical analysis in Section 7.

In another independent paper, Da and Xiu (2019) introduce a novel quasi maximum likelihood approach to estimate both the volatility and the autocovariances of moving-average microstructure noise. They also extend their estimators to general settings that allow for irregular observation times, intraday patterns of noise and jumps in asset prices. Their approach treats "large" and "small" microstructure noise in a uniform way which leads to a potential improvement in the convergence rate. Our approach is essentially of a nonparametric nature and provides unified estimators of a class of volatility functionals (see Theorem 4.1) including the asymptotic variance, which account for the feedback between finite sample and asymptotic biases. Our empirical study also has a different focus. Our investigation is not as extensive as in Da and Xiu (2019),[4] but we explicitly consider different sampling frequencies,[5] analyzing the autocovariance patterns of noise in connection to microstructure noise models and their impact on IV estimation.

The remainder of this paper is organized as follows. In Section 2, we introduce the basic setting and notation. In Section 3, we analyze realized volatility with dependent noise and develop consistent estimators of the second moments of noise. The pre-averaging method with dependent noise is studied

[4] Da and Xiu maintain a website to provide up-to-date daily annualized volatility estimates for all S&P 1500 index constituents, see http://dachxiu.chicagobooth.edu/#risklab.

[5] In their empirical studies, Da and Xiu (2019) only consider tick time sampling.

in Section 4. Section 5 introduces our multi-step estimators. Section 6 reports extensive simulation studies. Our empirical study is presented in Section 7. Section 8 concludes the paper. All proofs and some additional Monte Carlo simulation and empirical results are collected in an online appendix, see Li et al. (2019).

2 Framework and Assumptions

We state the following assumption regarding the efficient log-price process:

Assumption 2.1 (Efficient log-price)**.** *The efficient log-price process X follows a continuous Itô semi-martingale defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$:*

$$X_t = X_0 + \int_0^t b_s \mathrm{d}s + \int_0^t \sigma_s \mathrm{d}W_s, \tag{2}$$

where W is a standard Brownian motion, the drift process b_s is optional and locally bounded, and the volatility process σ_s is adapted with càdlàg paths.

We assume that all price observations are collected in the fixed time interval $[0, T]$, where without losing generality we let $T = 1$. We let $n+1$ be the number of observations and denote $\Delta_n = 1/n$. The observation times are given by $t_i^n = i\Delta_n, i = 0, \ldots, n$. We make the following assumption regarding the market microstructure noise:

Assumption 2.2 (Market microstructure noise)**.** *The noise process $(U_i)_{i\in\mathbb{N}}$ is defined on the probability space $(\Omega^{(0)}, \mathcal{G}, \mathbb{P}^{(0)})$, which has discrete filtrations $\mathcal{G}_i = \sigma(U_k : k \leq i)$, $\mathcal{G}^i = \sigma(U_k : k \geq i)$ that satisfy $\mathcal{G} = \mathcal{G}^\infty = \mathcal{G}_\infty$. Moreover, we assume*

1. *U is stationary and ρ-mixing and the mixing coefficients*[6] *$\{\rho_h\}_{h=1}^\infty$ decay at a polynomial rate, i.e., there exist some constants $C > 0, v > 0$ such that*

$$\rho_h \leq \frac{C}{h^v}. \tag{3}$$

2. *$v > 1$, $\mathbb{E}(U) = 0$, and all moments of noise exist.*

The mixing conditions in Assumption 2.2 item (1.) ensure that the noise process evaluated at different time instances, say, i and $i+h$, is increasingly limited in dependence when the lag h increases. In

[6]The *mixing coefficients* constitute a sequence satisfying

$$\rho_h = \sup\left\{|\mathbb{E}(V_k V_{k+h})| : \mathbb{E}(V_k) = \mathbb{E}(V_{k+h}) = 0,\ \|V_k\|_2 \leq 1,\ \|V_{k+h}\|_2 \leq 1,\ V_k \in \mathcal{G}_i,\ V_{k+h} \in \mathcal{G}^{i+h}\right\}.$$

We refer to Bradley (2007) or Chapter VIII of Jacod and Shiryaev (2003) for further details on and properties of mixing sequences.

particular, there exists some $C' > 0$ such that

$$|\gamma(h)| \leq \frac{C'}{h^v}, \tag{4}$$

where $\gamma(h) = \mathbf{Cov}(U_i, U_{i+h})$ is the autocovariance function of U, see Lemma VIII 3.102 in Jacod and Shiryaev (2003). We assume all moments of noise exist because this is required for the validity of Theorem 4.1 below for any even integer $r \geq 2$.

At stage n, we will denote U_i by U_i^n, $\forall i \leq n$. The i-th observed price is thus given by

$$Y_i^n = X_i^n + U_i^n, \tag{5}$$

where $X_i^n = X_{i\Delta_n}$.

Remark 2.1 (Microstructure noise in discrete time)**.** *We allow the noise process U to generate dependencies in* sampling time. *Hence, our noise process essentially constitutes a* discrete-time model *— it does not depend explicitly on the time between successive observations. Aït-Sahalia et al. (2005), Hansen and Lunde (2006), and Hansen et al. (2008) study various* continuous-time *models of dependent microstructure noise. In these continuous-time models, the noise component of a log-return over a time interval Δ is of order $O_p(\sqrt{\Delta})$, the same order as the logarithmic return of the efficient price. Our theory focuses primarily on sampling in* calendar time.[7] *In our simulations and empirical work, we also analyze sampling in* transaction time,[8] *and* tick time.[9]

Remark 2.2 (General dynamic properties of microstructure noise)**.** *Our assumptions on the dependence of noise are quite general, nesting many models as special cases including, for example, i.i.d. noise, q-dependent noise (i.e., $\gamma(h) = 0$, $\forall h > q$), ARMA(p, q) noise (see Mokkadem (1988)) and some long-memory processes (see Tsay (2005)). We note that AR(1) and AR(2) noise are studied in Barndorff-Nielsen et al. (2008) and Hendershott et al. (2013) respectively, q-dependent noise is considered by Hansen et al. (2008) and Hautsch and Podolskij (2013), while Gross-KlussMann and Hautsch (2013) study long-memory bid-ask spreads.*

Another recent strand of the literature explores the variety of microstructure data including observable information, seeking to parameterize the microstructure noise; see Li et al. (2016), Chaker (2017), Clinet and Potiron (2017) and Clinet and Potiron (2019). The parametrization allows for rich dynamics of

[7]Under this sampling scheme, Y_i^n (resp. X_i^n, U_i^n) is the observed log-price (resp. efficient log-price, microstructure noise) at regular time $i\Delta_n$, with $\Delta_n = 1/n$ in the main text.

[8]Under this sampling scheme, Y_i^n (resp. X_i^n, U_i^n) is the observed log-price (resp. efficient log-price, microstructure noise) associated with the i-th trade. The observation times $(t_i^n)_{0\leq i\leq n}$ can, in general, be deterministic or random, and regular or irregular.

[9]Tick time sampling removes all zero returns; see Aït-Sahalia et al. (2011) and Griffin and Oomen (2008). Hence, Y_i^n is by definition different from Y_{i-1}^n and Y_{i+1}^n under this sampling scheme.

the microstructure noise and at the same time improves the convergence rates of associated volatility estimators. Specifically, the noise component in these models can be serially correlated. The correlation is attributed to persistent observable quantities, e.g., trading volume and trading directions, that constitute the "observable part" of the microstructure noise. By contrast, we introduce an essentially nonparametric model of microstructure noise, without singling out the sources of the noise.

3 Estimation of the Variance and Covariances of Noise

In this section, we develop consistent estimators of the second moments of noise under Assumptions 2.1 and 2.2. These estimators will later serve as important inputs to adapt the pre-averaging method. We also analyze our estimators' finite sample properties.

3.1 Realized volatility with dependent noise

We start with the following preliminary result:

Proposition 3.1. *Assume that the efficient log-price satisfies Assumption 2.1, the observations follow (5), the noise process satisfies Assumption 2.2, and that $\mathcal{G}$ is independent of $\mathcal{F}$. Let $j > 0$ be a fixed integer and assume the sequence of integers j_n satisfies $j_n \to \infty$, $j_n \Delta_n \to 0$. Then we have the following convergences in probability as $n \to \infty$:*

$$\widehat{\langle Y, Y \rangle}(j)_n := \frac{\sum_{i=0}^{n-j} (Y_{i+j}^n - Y_i^n)^2}{2(n-j+1)} \xrightarrow{\mathbb{P}} \gamma(0) - \gamma(j), \tag{6}$$

$$\widehat{\gamma(0)}_n := \frac{\sum_{i=0}^{n-j_n} (Y_{i+j_n}^n - Y_i^n)^2}{2(n-j_n+1)} \xrightarrow{\mathbb{P}} \gamma(0), \tag{7}$$

$$\widehat{\gamma(j)}_n := \widehat{\gamma(0)}_n - \widehat{\langle Y, Y \rangle}(j)_n \xrightarrow{\mathbb{P}} \gamma(j). \tag{8}$$

The special case of (6) that occurs when $j = 1$ appears in Aït-Sahalia et al. (2011) assuming exponential decay. We also note that in the most recent version of Jacod et al. (2017) similar estimators as $\widehat{\langle Y, Y \rangle}(j)_n$ are mentioned but without formal analysis of their limiting behavior. To our best knowledge, our paper is the first to estimate the variance and covariances of noise using realized volatility under a general dependent noise setting.

3.2 Finite sample bias correction

The theoretical validity of our realized volatility estimators in (6)–(8) hinges on the increasing availability of observations in a fixed time interval, the so-called *infill asymptotics*. In general, an estimator derived from asymptotic results can, however, behave very differently in finite samples. Our realized volatility estimators of the second moments of noise are an example for which the asymptotic theory provides a poor representation of the estimators' finite sample behavior.[10]

Intuitively, the finite sample bias stems from the diffusion component, when computing the realized volatility $\widehat{\langle Y,Y \rangle}(j)_n$ over large lags j in a finite sample, and we will explain later (e.g., in Remark 3.3) why it is critically relevant to account for it in real applications. In the remainder of this section, we assume the drift b_t in (2) to be zero. As shown by, for example, Bandi and Russell (2008) and Lee and Mykland (2012) this is not restrictive in high-frequency analysis. This will be confirmed in our Monte Carlo simulation studies in Section 6 and Appendix B.

Proposition 3.2. *Assume that the efficient log-price follows (2) with $b_s = 0\ \forall s$, and assume there is some $\delta > 0$ so that σ_t is bounded for all $t \in [0,\delta] \cup [1-\delta, 1]$. Furthermore, assume the observations follow (5), the noise process satisfies Assumption 2.2 and $\mathcal{G}$ is independent of $\mathcal{F}$. Then,*

$$\mathbb{E}_\sigma\left(\widehat{\langle Y,Y \rangle}(j)_n\right) = \frac{j\mathrm{IV}}{2(n-j+1)} + \gamma(0) - \gamma(j) + O_p\left(j^2/n^2\right), \tag{9}$$

where $\mathrm{IV} := \int_0^1 \sigma_t^2 \mathrm{d}t$ is the integrated volatility. Here, $\mathbb{E}_\sigma(\cdot)$ denotes the expectation conditional on the volatility path.

Remark 3.1. *If σ_t is locally bounded, then the assumptions on σ_t required for Proposition 3.2 will hold. The regularity conditions with respect to σ_t in Proposition 3.2 trivially hold if the volatility process is assumed to be continuous. (Volatility is usually assumed to be continuous when making finite sample bias corrections.)*

Remark 3.2. *Let $j = 1$. In that special case the result in Proposition 3.2 bears similarities to Theorem 1 in Hansen and Lunde (2006). Contrary to Hansen and Lunde (2006) we assume that the efficient log-price X is independent of the noise U. Therefore, any correlations between the two drop out.*

Proposition 3.2 reveals that $\widehat{\langle Y,Y \rangle}(j)_n - \frac{j\mathrm{IV}}{2(n-j+1)}$ will be a better estimator of $\gamma(0) - \gamma(j)$ in finite samples than $\widehat{\langle Y,Y \rangle}(j)_n$, and this motivates the following finite sample bias corrected estimators:

$$\widehat{\langle Y,Y \rangle}^{(\mathrm{adj})}(j)_n := \widehat{\langle Y,Y \rangle}(j)_n - \frac{\hat{\sigma}^2 j}{2(n-j+1)}; \tag{10}$$

$$\widehat{\gamma(0)}_n^{(\mathrm{adj})} := \widehat{\gamma(0)}_n - \frac{\hat{\sigma}^2 j_n}{2(n-j_n+1)}; \tag{11}$$

$$\widehat{\gamma(j)}_n^{(\mathrm{adj})} := \widehat{\gamma(0)}_n^{(\mathrm{adj})} - \widehat{\langle Y,Y \rangle}^{(\mathrm{adj})}(j)_n; \tag{12}$$

[10] This applies to the local averaging estimators developed in Jacod et al. (2017) as well; see Footnote 13 for further details.

where $\hat{\sigma}^2$ is an estimator of IV. We note that the bias corrected estimators are still consistent, as the fraction $\frac{j}{n-j+1}$ is negligible when j is much smaller than n.

Remark 3.3 (Why the finite sample bias matters). *We now explain why the finite sample bias correction is crucial in applications. We first rewrite* (9)*:*

$$\begin{aligned}\mathbb{E}_\sigma\left(\widehat{\langle Y,Y\rangle}(j)_n\right) &= \frac{j\mathrm{IV}}{2(n-j+1)} + \gamma(0) - \gamma(j) + O_p\left(j^2/n^2\right) \\ &= (\gamma(0)-\gamma(j))\left(1 + \frac{\frac{j}{2(n-j+1)}}{\frac{\gamma(0)-\gamma(j)}{\mathrm{IV}}}\right) + O_p\left(j^2/n^2\right).\end{aligned} \tag{13}$$

Observe that the finite sample bias is determined by the ratio of the two terms $\frac{j}{2(n-j+1)}$ *and* $\frac{\gamma(0)-\gamma(j)}{\mathrm{IV}}$*. The first term,* $\frac{j}{2(n-j+1)}$*, depends on the data frequency* (n) *and "target parameters"* (j)*; the second term,* $\frac{\gamma(0)-\gamma(j)}{\mathrm{IV}}$*, is the (latent) noise-to-signal ratio. If the second term is "relatively larger (smaller)" than the first one, then the finite sample bias will be small (large). In other words, the finite sample bias is not only determined by the data frequency and target parameters, but also by other properties of the underlying efficient price and noise processes.*

In high-frequency financial data, the noise-to-signal ratio $\frac{\gamma(0)}{\mathrm{IV}}$ *is typically small, but it can vary from* $O(10^{-2})$ *(see Bandi and Russell (2006)) to* $O(10^{-6})$ *(see Christensen et al. (2014)) in empirical studies. The ratio* $\frac{j}{2(n-j+1)}$*, while typically small as well, can still be* relatively *large, depending on the specific situation. Consider the following two scenarios:*

1) We have ultra high-frequency data with $n = O(10^5)$ *(recall that the number of seconds in a business day is 23,400), and we select* $j_n = 20$*. Then, the ratio* $\frac{j_n}{2(n-j_n+1)} = O(10^{-4})$*.*

2) We have i.i.d. noise and we would like to estimate the variance of noise by $\widehat{\langle Y,Y\rangle}(1)_n$ *using high-frequency data with average duration of 20 seconds (thus* $n \approx 10^3$*); see, e.g., Bandi and Russell (2006). Hence,* $\frac{j}{2(n-j+1)} = O(10^{-3})$*.*

In both scenarios, the ratio of $\frac{j}{2(n-j+1)}$ *and* $\frac{\gamma(0)-\gamma(j)}{\mathrm{IV}}$ *can vary widely, depending on the magnitude of the latent noise-to-signal ratio. It is then clear from the first line of* (13) *that the finite sample bias term, which is proportional to the IV, may well wipe out the variance of noise, depending on the specific situation.*

Remark 3.4. *Note that increasing the sample size by extending the time horizon to* $[0, T]$ *with large* T *will not remove the finite sample bias. Hence, the finite sample bias may be viewed as a* low frequency bias*.*

Throughout the remainder of this paper, we assume the following conditions hold:[11]

$$v > 3, \quad j_n \asymp \Delta_n^{-\delta}, \quad \ell_n \asymp \Delta_n^{-\kappa}, \quad \delta \in \left(\frac{5}{36}, \frac{1}{6}\right), \kappa \in \left(\frac{1}{8}, \frac{1}{6}\right), \tag{14}$$

with ℓ_n another sequence of integers. The following proposition provides an estimator of the "long-run variance" of microstructure noise. As we shall see later, the long-run variance of noise appears as an asymptotic bias in the de-noise method developed in this paper.

Proposition 3.3. *Assume that the efficient log-price satisfies Assumption 2.1, the observations follow (5), the noise process satisfies Assumption 2.2 and $\mathcal{G}$ is independent of $\mathcal{F}$. Define*

$$\widehat{\Sigma}_{U_n} := \widehat{\gamma(0)}_n + 2\sum_{j=1}^{\ell_n} \widehat{\gamma(j)}_n, \tag{15}$$

where $\widehat{\gamma(0)}_n$ and $\widehat{\gamma(j)}_n$ are defined in (7) and (8). Then,

$$\widehat{\Sigma}_{U_n} \xrightarrow{\mathbb{P}} \Sigma_U, \tag{16}$$

where

$$\Sigma_U = \gamma(0) + 2\sum_{j=1}^{\infty} \gamma(j). \tag{17}$$

For i.i.d. noise, Σ_U reduces to $\gamma(0)$, and it is known (see Zhang et al. (2005) and Bandi and Russell (2008)) that the variance of noise can then be consistently estimated by the standardized realized volatility of observed returns. However, when noise is dependent we face a much more complex situation: all variance and covariance terms constitute Σ_U. Nevertheless, Proposition 3.3 above provides a consistent estimator of Σ_U.

4 The Pre-Averaging Method with Dependent Noise

In this section, we adapt a popular "de-noise" method — the pre-averaging method — to allow for serially dependent noise in our general setting. The pre-averaging method was originally introduced by Podolskij and Vetter (2009b) (see also Jacod et al. (2009), Jacod et al. (2010), Podolskij and Vetter (2009a), Hautsch and Podolskij (2013), and the textbook treatment in Aït-Sahalia and Jacod (2014)). We first construct our pre-averaged statistics based on non-overlapping sampling blocks and next based

[11] Some results, e.g., Proposition 3.3, hold already under weaker conditions. The conditions (14) are, however, needed to establish our main theorems in the next sections.

on overlapping sampling blocks, in both cases using general weight functions.

4.1 Pre-averaging based on non-overlapping intervals

Let k_n be a sequence of integers, with $k_n \to \infty$ and $k_n \Delta_n \to 0$ as $n \to \infty$, satisfying

$$\sqrt{\Delta_n} k_n = \theta + o(\Delta_n^{1/4}), \tag{18}$$

where $\theta > 0$ is a constant. Furthermore, let g be a general kernel (i.e., weight function). We assume g is continuous, piecewise C^1 with a piecewise Lipschitz derivative g', and satisfies $g(s) = 0$, $\forall s \notin (0, 1)$, and $\int_0^1 g^2(s) \mathrm{d}s > 0$, as in Jacod et al. (2009). We introduce the following notation associated with g:

$$\begin{cases} g_i^n = g(i/k_n); \quad \overline{g}_i^n = g_{i+1}^n - g_i^n; \\ \phi_0^n = \frac{1}{k_n} \sum_{i \in \mathbb{Z}} \left(g_i^n\right)^2; \quad \phi_1^n(j) = k_n \sum_{i \in \mathbb{Z}} \overline{g}_i^n \overline{g}_{i-j}^n; \\ \phi_0(s) = \int_s^1 g(u) g(u-s) \mathrm{d}u; \quad \phi_1(s) = \int_s^1 g'(u) g'(u-s) \mathrm{d}u; \\ \Phi_{ij} = \int_0^1 \phi_i(s) \phi_j(s) \mathrm{d}s, \ \psi_i = \phi_i(0), \ \ i, j \in \{0, 1\}. \end{cases}$$

Example 4.1 (Triangular kernel)**.** *A simple canonical example of g is given by the* triangular kernel $g(x) = x \wedge (1 - x)$. *Then,*

$$\psi_0 = 1/12, \quad \psi_1 = 1, \quad \Phi_{00} = 151/80640, \quad \Phi_{01} = 1/96, \quad \Phi_{11} = 1/6.$$

For any sequence $\{Z_i^n\}_{i=0}^n$, denote $\Delta_i^n Z = Z_i^n - Z_{i-1}^n$, $i = 1, 2, \ldots$, and let its pre-averaged value be given by

$$\overline{Z}_i^n := \sum_{j=1}^{k_n - 1} g_j^n \Delta_{i+j}^n Z = -\sum_{j=0}^{k_n - 1} \overline{g}_j^n Z_{i+j}^n, \qquad i = 0, 1, \ldots. \tag{19}$$

Furthermore, let $M_n = \lfloor \frac{n}{k_n} \rfloor$, where $\lfloor \cdot \rfloor$ is the floor function. For any real $r \geq 2$, the pre-averaged statistics of the log-price process Y based on *non-overlapping intervals* are defined as follows:

$$\mathrm{PAV}(Y, r)_n := n^{\frac{r-2}{4}} \sum_{m=0}^{M_n - 1} \left| \overline{Y}_{m k_n}^n \right|^r, \quad r \geq 2. \tag{20}$$

Under our general setting of dependent noise, we establish in the following results first a consistency theorem for the general functional form of the pre-averaged statistics, based on which we derive a consistent estimator of the IV, and next a central limit theorem providing the associated limit distribution, with a consistent estimator of the asymptotic variance.

Theorem 4.1. *Assume that the efficient log-price satisfies Assumption 2.1, the observations follow (5), and the noise process satisfies Assumption 2.2. Furthermore, assume $\mathcal{G}$ and $\mathcal{F}$ are independent. Then, for any even integer $r \geq 2$,*

$$\mathrm{PAV}(Y,r)_n \xrightarrow{\mathbb{P}} \mathrm{PAV}(Y,r) := \frac{\mu_r}{\theta} \int_0^1 \left(\theta \psi_0 \sigma_s^2 + \frac{\psi_1}{\theta} \Sigma_U \right)^{\frac{r}{2}} \mathrm{d}s, \tag{21}$$

where Σ_U is defined in (17) and $\mu_r = \mathbb{E}(Z^r)$ for a standard normal random variable Z.

Aided by this result, we obtain consistent estimators of the IV and the integrated quarticity IQ $:= \int_0^1 \sigma_s^4 \mathrm{d}s$, as follows:

Corollary 4.1. *Under the assumptions of Theorem 4.1, we have the following consistency result for the IV and the IQ:*

$$\widehat{\mathrm{IV}}_n := \frac{\mathrm{PAV}(Y,2)_n}{\psi_0} - \frac{\psi_1 \widehat{\Sigma}_{U_n}}{\psi_0 \theta^2} \xrightarrow{\mathbb{P}} \mathrm{IV}, \tag{22}$$

$$\widehat{\mathrm{IQ}}_n := \frac{\mathrm{PAV}(Y,4)_n}{3\psi_0^2 \theta} - \frac{2\psi_1 \widehat{\Sigma}_{U_n} \widehat{\mathrm{IV}}_n}{\psi_0 \theta^2} - \frac{\psi_1^2 \left(\widehat{\Sigma}_{U_n} \right)^2}{\theta^4 \psi_0^2} \xrightarrow{\mathbb{P}} \mathrm{IQ}, \tag{23}$$

where $\widehat{\Sigma}_{U_n}$ is defined in (15).

Theorem 4.2. *Assume all conditions in Theorem 4.1 hold. Furthermore, assume that the process σ is a continuous Itô semimartingale. Then,*

$$\Delta_n^{-\frac{1}{4}} \left(\widehat{\mathrm{IV}}_n - \mathrm{IV} \right) \xrightarrow{\mathcal{L}-s} \sqrt{\frac{2}{\theta \psi_0^2}} \int_0^1 \left(\theta \psi_0 \sigma_s^2 + \frac{\psi_1}{\theta} \Sigma_U \right) \mathrm{d}W_s', \tag{24}$$

where $\xrightarrow{\mathcal{L}-s}$ denotes stable convergence in law and where W' is a standard Wiener process independent of $\mathcal{F}$. Moreover, letting $\widehat{\Sigma}_n := 2\mathrm{PAV}(Y,4)_n / 3\psi_0^2$, we have that $\Delta_n^{-\frac{1}{4}} \left(\widehat{\mathrm{IV}}_n - \mathrm{IV} \right) / \sqrt{\widehat{\Sigma}_n}$ converges stably in law to a standard normal random variable, which is independent of $\mathcal{F}$.

A main advantage of the pre-averaging approach and the associated estimators introduced in this section is their simplicity. In fact, we obtain from Theorem 4.1 a class of consistent estimators of $\int_0^1 \sigma_s^r \mathrm{d}s$ with arbitrary even integer r, since we have a consistent estimator of Σ_U. When only estimation of the IV is concerned, this leads to a simple estimator of the asymptotic variance of the IV estimator.[12]

[12] Our simulation experiments presented later show that, compared to the pre-averaging estimators based on overlapping intervals introduced in the next subsection, the pre-averaging estimators based on non-overlapping intervals often deliver a somewhat smaller bias, although their standard deviations are typically somewhat larger.

4.2 Pre-averaging based on overlapping intervals

Now we extend our pre-averaging estimator of the IV in two directions. First, we allow for *overlapping intervals* to conduct pre-averaging; second, we accommodate more general stochastic volatility processes when deriving the respective limit distribution. (We recall that we assumed the process σ to be a continuous Itô semimartingale in Theorem 4.2.)

In particular, we propose the following estimator of the IV, with $\widehat{\Sigma}_{U_n}$ as introduced in (15):

$$\widetilde{\mathrm{IV}}_n := \frac{\sqrt{\Delta_n}}{\theta\psi_0} \sum_{i=0}^{n-k_n+1} \left(\overline{Y}_i^n\right)^2 - \frac{\psi_1 \widehat{\Sigma}_{U_n}}{\theta^2\psi_0}. \tag{25}$$

Theorem 4.3. *Assume that the efficient log-price satisfies Assumption 2.1, the observations follow (5), the noise process satisfies Assumption 2.2, and $\mathcal{G}$ is independent of $\mathcal{F}$. Then,*

$$\Delta_n^{-\frac{1}{4}} \left(\widetilde{\mathrm{IV}}_n - \mathrm{IV}\right) \xrightarrow{\mathcal{L}-s} \Upsilon_1, \tag{26}$$

with $\Upsilon_t = \int_0^t V_s \mathrm{d}W'_s$, where W' is a standard Wiener process independent of $\mathcal{F}$, and where V_t satisfies

$$V_t^2 := \frac{4}{\psi_0^2} \left(\Phi_{00}\theta\sigma_t^4 + 2\Phi_{01}\frac{\sigma_t^2\Sigma_U}{\theta} + \frac{\Phi_{11}\Sigma_U^2}{\theta^3} \right). \tag{27}$$

Remark 4.1. *The tuning parameter θ (recall (18)) can be chosen such that it minimizes the asymptotic variance, which will improve the efficiency of our estimators. The optimal θ is given by*

$$\theta^* = \left(\frac{\sqrt{\Phi_{01}^2 \mathrm{IV}^2 \Sigma_U^2 + 3\Phi_{00}\Phi_{11}\Sigma_U^2 \mathrm{IQ}} + \Phi_{01}\mathrm{IV}\Sigma_U}{\Phi_{00}\mathrm{IQ}} \right)^{1/2}. \tag{28}$$

The optimal choice of θ requires an estimate of IQ. *Therefore, we provide a consistent estimator, as follows:*

$$\widetilde{\mathrm{IQ}}_n := \frac{\sum_{i=0}^{n-k_n+1}\left(\overline{Y}_i^n\right)^4}{3\theta^2\psi_0^2} - \frac{2\psi_1\widehat{\Sigma}_{U_n}\widetilde{\mathrm{IV}}_n}{\psi_0\theta^2} - \frac{\psi_1^2\left(\widehat{\Sigma}_{U_n}\right)^2}{\theta^4\psi_0^2} \xrightarrow{\mathbb{P}} \mathrm{IQ}. \tag{29}$$

Note that that $\widetilde{\mathrm{IQ}}_n$ is analogous to $\widehat{\mathrm{IQ}}_n$ introduced in (23).

To apply the limit distribution result in Theorem 4.3 above to construct confidence intervals, we need a consistent estimator of the asymptotic variance $\int_0^1 V_t^2 \mathrm{d}t$. Among other possibilities, we propose the following:

$$\widetilde{\Sigma}_n := \frac{4\Phi_{00}}{3\theta\psi_0^4}\sum_{i=0}^{n-k_n+1}\left(\overline{Y}_i^n\right)^4 + \frac{8\widehat{\Sigma}_{U_n}\widetilde{\mathrm{IV}}_n}{\theta\psi_0^2}\left(\Phi_{01} - \frac{\psi_1\Phi_{00}}{\psi_0}\right) + \frac{4\left(\widehat{\Sigma}_{U_n}\right)^2 T}{\theta^3\psi_0^2}\left(\Phi_{11} - \frac{\psi_1^2\Phi_{00}}{\psi_0^2}\right). \tag{30}$$

Corollary 4.2. *Under the assumptions of Theorem 4.3, we have*

$$\widetilde{\Sigma}_n \xrightarrow{\mathbb{P}} \int_0^1 V_t^2 \mathrm{d}t. \tag{31}$$

Therefore, the sequence $\Delta_n^{-\frac{1}{4}}\left(\widetilde{\mathrm{IV}}_n - \mathrm{IV}\right)/\sqrt{\widetilde{\Sigma}_n}$ converges stably in law to a standard normal random variable, which is independent of $\mathcal{F}$.

Remark 4.2 (Irregular Observation Schemes)**.** *We note that, following similar arguments as in Jacod and Mykland (2015), our results, in particular those in Theorem 4.3, extend to (i.e., are robust to) mildly irregular observation schemes, as follows. Let $\mathcal{T}$ be a function with strictly positive Lipschitz derivative. Assume $\mathcal{T}(0) = 0$ and $\mathcal{T}(1) = 1$. Now let $\tilde{t}_i^n := \mathcal{T}(i\Delta_n)$. Such irregular observation schemes have been considered e.g., by Barndorff-Nielsen et al. (2008) and Mykland and Zhang (2012).*

First, we note that such a time transformation theoretically does not *affect the microstructure noise process, as the noise is a discrete-time process that does not depend on the time between successive observations. Thus, under the new observation scheme, we have that*

$$Y_{\tilde{t}_i^n} = X_{\tilde{t}_i^n} + U_i^n. \tag{32}$$

Denote the time-transformed efficient price process by $X_{\mathcal{T},t} := X_{\mathcal{T}(t)}$ with $b_{\mathcal{T},t} := b_{\mathcal{T}(t)}\mathcal{T}'(t)$ and $\sigma_{\mathcal{T},t} := \sigma_{\mathcal{T}(t)}\sqrt{\mathcal{T}'(t)}$.

Several conclusions are immediate. First, the new process $X_{\mathcal{T}}$ satisfies Assumption 2.1; second, under the transformation $\mathcal{T}$, the irregular observation scheme becomes regular in the sense that $X_{\tilde{t}_i^n} = X_{\mathcal{T},i\Delta_n}$; third, the integrated volatility and the integrated quarticity are unchanged due to the properties of $\mathcal{T}$, upon a change of variable; finally, the probabilistic and statistical behavior of the noise is unchanged, in particular, Σ_U is unchanged and its consistent estimator remains valid.

Thus, upon replacing $i\Delta_n$ by $\tilde{t}_i^n$, we can apply our Theorem 4.3 to observed noisy prices $Y_{\mathcal{T},}{}_i^n = X_{\mathcal{T},}{}_i^n + U_i^n$, which agrees exactly with (32). The limit distribution remains valid, in particular, the asymptotic variance of the IV estimator remains the same.

Remark 4.3 (Jumps in the Efficient Price)**.** *Assumption 2.1 does not allow for jumps in the efficient price process X. (Jumps in the stochastic volatility process are allowed.) We note from the proof of Proposition 3.1 that jumps in the efficient price will not affect the convergences of the realized volatility* *estimators of the second moments of noise, as the noise has larger asymptotic orders. For the pre-averaging estimators, we conjecture that under suitable conditions both $\widehat{\mathrm{IV}}_n$ and $\widetilde{\mathrm{IV}}_n$ will converge to the quadratic variation of X instead of to the IV. One can apply the truncation method (Mancini (2001)) to eliminate the jumps. But this is beyond the scope of this paper. In this context, it is worth mentioning an extensive empirical study by Christensen et al. (2014), in which the authors show that, as far as IV estimation is concerned, the jump component of the efficient price process in (very) high-frequency data typically only accounts for a small portion of the total price variation.*

4.3 Efficiency

It is well-known that estimators of volatility from noisy observations can achieve efficiency when the volatility is a constant, c_σ, i.e., the integrated volatility over $[0, t]$ equals tc_σ, and the noise takes the form of Gaussian white (i.e., i.i.d.) noise with variance $\mathbf{Var}(U)$; see Gloter and Jacod (2001a) and Gloter and Jacod (2001b) for a detailed account. In this case, an efficient estimator of the IV will converge at rate $\Delta_n^{-\frac{1}{4}}$ with an asymptotic variance equal to $\Sigma_t^{\mathrm{opt}} = 8tc_\sigma^{3/2}\sqrt{\mathbf{Var}(U)}$. When the assumption of constant volatility is maintained but the noise is serially dependent, the optimal asymptotic variance becomes $\overline{\Sigma}_t^{\mathrm{opt}} = 8tc_\sigma^{3/2}\sqrt{\Sigma_U}$, with the variance of noise replaced by the long-run variance of noise; see Da and Xiu (2019). We can show that the asymptotic variance of our estimator $\widetilde{\mathrm{IV}}_n$, with the optimally selected θ (recall Remark 4.1) and using the triangular kernel, is quite close to $\overline{\Sigma}_t^{\mathrm{opt}}$ under constant volatility:

$$\frac{\int_0^t V_s^2 \mathrm{d}s}{\overline{\Sigma}_t^{\mathrm{opt}}} \approx 1.07. \tag{33}$$

With stochastic volatility, it is still possible to achieve (33) asymptotically using local estimation — divide all observations into B blocks and perform estimation on each block and then aggregate the block estimates; see, e.g., Jacod and Mykland (2015), Clinet and Potiron (2018) and Da and Xiu (2019). Our simulation experience (not reported here) shows that in finite samples our estimators often do, but need not always, improve when following such a local estimation procedure. In those cases in which there is a lack of improvement, this may be partially due to a relatively worse estimation of the optimal θ in a smaller sample.

Any proper estimation of θ, whether local or global, requires accurate estimates of characteristics of the efficient price and noise processes. We will show through our extensive simulations and empirical studies that model (mis)specification and finite sample biases play first-order roles in the estimation of such characteristics, and that our multi-step method introduced in the next section provides a robust approach. In our analyses presented later, we don't pursue local estimation, but focus on illustrating the robustness of our multi-step approach to model misspecification and to finite sample biases.

5 Multi-Step Estimators

In this section, we introduce our multi-step estimators of the IV and the second moments of noise based on both our asymptotic theory and finite sample analysis.

We observe from Theorem 4.3 that the second moments of noise contribute to an *asymptotic bias* in the estimation of the IV. Our finite sample analysis indicates, however, that we need an estimator of the IV to correct the *finite sample bias* when estimating the second moments of noise. Our multi-step estimators are specifically designed for the purpose of correcting the "interlocked" bias.

In the first step, we ignore the dependence in noise and estimate the variance of noise by realized volatility. Hence, our first-step estimators of the second moments of noise are given by

$$\widetilde{\gamma(0)}_n^{(1)} := \widehat{\langle Y, Y\rangle}(1)_n; \quad \widetilde{\gamma(j)}_n^{(1)} := 0,\ j \neq 0; \quad \widetilde{\Sigma}_{U_n}^{(1)} := \widetilde{\gamma(0)}_n^{(1)}. \tag{34}$$

Next, we proceed with the pre-averaging method to obtain the first-step estimator of the IV (cf. (25)):

$$\widetilde{\mathrm{IV}}_n^{(1)} = \frac{\sqrt{\Delta_n}}{\theta\psi_0}\sum_{i=0}^{n-k_n+1}\left(\overline{Y}_i^n\right)^2 - \frac{\psi_1\widetilde{\Sigma}_{U_n}^{(1)}}{\theta^2\psi_0}. \tag{35}$$

To initiate the second step, we first replace $\hat{\sigma}^2$ by $\widetilde{\mathrm{IV}}_n^{(1)}$ in (10) and (11) and obtain the second-step estimators of the variance and covariances of noise. They will in turn be used to correct the asymptotic bias in the estimation of the IV, to eventually obtain the second-step estimator of the IV. Upon iterating this procedure, one obtains multi-step estimators. Specifically, for any $k \geq 2$, we define the k-step estimators recursively as follows:

$$\widetilde{\langle Y, Y\rangle}(j)_n^{(k)} := \widehat{\langle Y, Y\rangle}(j)_n - \frac{j\widetilde{\mathrm{IV}}_n^{(k-1)}}{2(n-j+1)}; \tag{36}$$

$$\widetilde{\gamma(0)}_n^{(k)} := \widehat{\gamma(0)}_n - \frac{j_n\widetilde{\mathrm{IV}}_n^{(k-1)}}{2(n-j_n+1)}; \tag{37}$$

$$\widetilde{\gamma(j)}_n^{(k)} := \widetilde{\gamma(0)}_n^{(k)} - \widetilde{\langle Y, Y\rangle}(j)_n^{(k)}; \tag{38}$$

$$\widetilde{\Sigma}_{U_n}^{(k)} := \widetilde{\gamma(0)}_n^{(k)} + 2\sum_{j=1}^{\ell_n}\widetilde{\gamma(j)}_n^{(k)}; \tag{39}$$

$$\widetilde{\mathrm{IV}}_n^{(k)} := \frac{\sqrt{\Delta_n}}{\theta\psi_0}\sum_{i=0}^{n-k_n+1}\left(\overline{Y}_i^n\right)^2 - \frac{\psi_1\widetilde{\Sigma}_{U_n}^{(k)}}{\theta^2\psi_0}; \tag{40}$$

$$\widetilde{\Sigma}_{\mathrm{IV}_n}^{(k)} := \frac{4\Phi_{00}}{3\theta\psi_0^4}\sum_{i=0}^{n-k_n+1}\left(\overline{Y}_i^n\right)^4 + \frac{8\widetilde{\Sigma}_{U_n}^{(k)}\widetilde{\mathrm{IV}}_n^{(k)}}{\theta\psi_0^2}\left(\Phi_{01} - \frac{\psi_1\Phi_{00}}{\psi_0}\right) + \frac{4\left(\widetilde{\Sigma}_{U_n}^{(k)}\right)^2}{\theta^3\psi_0^2}\left(\Phi_{11} - \frac{\psi_1^2\Phi_{00}}{\psi_0^2}\right). \tag{41}$$

We state the following theorem:

Theorem 5.1. *Under the assumptions of Theorem 4.3, for any fixed $K \in \mathbb{N}^*$, we have*

$$\widetilde{\Sigma}_{U_n}^{(K)} \xrightarrow{\mathbb{P}} \Sigma_U, \tag{42}$$

and the sequence $\Delta_n^{-\frac{1}{4}}\left(\widetilde{\mathrm{IV}}_n^{(K)} - \mathrm{IV}\right) / \sqrt{\widetilde{\Sigma}_{\mathrm{IV}_n}^{(K)}}$ converges stably in law to a standard normal random variable, which is independent of $\mathcal{F}$.

We note that, for brevity, our multi-step estimators introduced above are based only on the pre-averaging estimators using overlapping intervals. Of course, we can adopt the same approach and develop, by analogy, consistent and asymptotically normal multi-step estimators from the pre-averaging estimators using non-overlapping intervals as well. They will henceforth be denoted by $\widehat{\mathrm{IV}}_n^{(k)}$ and will be analyzed alongside $\widetilde{\mathrm{IV}}_n^{(k)}$ later.

Remark 5.1. *As the simulation results in the next section will reveal, our multi-step estimators introduced above perform well. An advantage of our multi-step estimators is that they are quite robust to the choice of the tuning parameter θ. To offer some insight into this issue, we briefly analyze the relationship between the choice of θ and the theoretical finite sample bias of two estimators: our $\widetilde{\mathrm{IV}}_n$ and the benchmark estimator $\widetilde{\mathrm{IV}}_n^{\mathrm{JLZ}}$ recently proposed by Jacod et al. (2019), which employs the local averaging (LA) method to correct the asymptotic bias of pre-averaging estimators. A simple calculation shows that the finite sample errors of $\widetilde{\mathrm{IV}}_n$ and $\widetilde{\mathrm{IV}}_n^{\mathrm{JLZ}}$ (as a percentage) are approximately given by*

$$\mathrm{Err}_{\mathrm{RV}} \approx \frac{(2\ell_n + 1)j_n + \sum_{|\ell| \le \ell_n} |\ell|\, \phi_1^n(\ell)}{2n\theta^2\psi_0}, \quad \mathrm{Err}_{\mathrm{JLZ}} \approx \frac{4K_n \sum_{|\ell| \le \ell_n} \phi_1^n(\ell)}{3n\theta^2\psi_0}, \tag{43}$$

respectively, where K_n is the tuning parameter of the LA method. While these errors can be significant for both estimators, and moreover a small change in θ can lead to sharp changes in the errors, our multi-step estimators are specifically designed to remove this error. Consequently, they are much more robust to changes in θ than estimators without unified bias corrections.

6 Simulation Study

6.1 Simulation design

Motivated by the empirical studies in Aït-Sahalia et al. (2011), we consider an ARMA(1,1) noise process U given by the following dynamics:

$$U_t = e_t + \epsilon_t, \tag{44}$$

where e is centered i.i.d. Gaussian and ϵ is an AR(1) process with first-order coefficient ι, $|\iota| < 1$. We will examine the performance of our estimators for different values of this coefficient: $\iota \in \{-0.7, -0.3, 0, 0.3, 0.7\}$. The processes e and ϵ are assumed to be statistically independent. We set $\mathbb{E}(e_t^2) = 1.9 \times 10^{-7}$, and $\mathbb{E}(\epsilon_t^2) = 1.3 \times 10^{-7}$. These values are chosen to mimic the results of our empirical studies.

We assume that the efficient price is generated by the following dynamics:

$$\begin{cases} \mathrm{d}X_t = -\kappa_1(X_t - \mu_1)\mathrm{d}t + \sigma_t \mathrm{d}W_t, \\ \mathrm{d}\sigma_t^2 = \kappa_2\left(\mu_2 - \sigma_t^2\right)\mathrm{d}t + \kappa_3\sigma_t \mathrm{d}B_t + \xi_t \mathrm{d}N_t, \end{cases}$$

where B and W are standard Brownian motions with quadratic covariation $\langle B, W\rangle_t = \varrho t$, N is a Poisson process with parameter λ, and ξ_t is an independent jump size following an exponential distribution with parameter κ_3. We set the parameters as follows: $\kappa_1 = 0.5$, $\mu_1 = 1.6$, $\kappa_2 = 5/252$, $\mu_2 = 0.04/252$, $\kappa_3 = 0.05/252$, $\lambda = 3$, and $\varrho = -0.5$. We assume the processes X and U to be mutually independent. We simulate each sample path within a fixed time interval $[0, 1]$ that represents one trading day.

6.2 Realized volatility estimators of the second moments of noise

To get a first impression of the properties of our estimator $\widehat{\langle Y, Y\rangle}(j)_n$ defined in (6), we plot $\widehat{\langle Y, Y\rangle}(j)_n$ against the number of lags j in Figure 2. In addition to $\widehat{\langle Y, Y\rangle}(j)_n$, we also plot the bias adjusted version $\widehat{\langle Y, Y\rangle}^{(\mathrm{adj})}(j)_n$ defined in (10), in which we employ three "approximations" to the IV that $\widehat{\langle Y, Y\rangle}^{(\mathrm{adj})}(j)_n$ depends on: $\hat{\sigma}_H^2 = 1.2\mathrm{IV}$, $\hat{\sigma}_M^2 = \mathrm{IV}$, and $\hat{\sigma}_L^2 = 0.8\mathrm{IV}$. Figure 2 shows that a prominent feature of our realized volatility estimator $\widehat{\langle Y, Y\rangle}(j)_n$ is that it deviates from its stochastic limit $\gamma(0) - \gamma(j)$ almost linearly in the number of lags j, as predicted by Proposition 3.2. The deviation, induced by the finite sample bias, can be largely corrected when only rough "estimates" of the IV are available. In the ideal but infeasible situation that we know exactly the true volatility ($\hat{\sigma}_M^2 = \mathrm{IV}$), the bias corrected estimators almost perfectly match the underlying true values.

Next, we estimate the second moments of noise by our realized volatility estimators (RV) and, for comparison purposes, by the local averaging estimators (LA) proposed by Jacod et al. (2017). We demonstrate the importance of the finite sample bias correction to obtain accurate estimates, and this applies to both estimators.[13] In Figure 3, we plot the means of the autocorrelations of noise estimated

[13] The finite sample bias corrected local averaging estimators of the noise covariances are given by

$$\widehat{R}(j)_n = \frac{1}{n}U((0,j))_n - \frac{K_n}{n}\left(\frac{4}{3}\hat{\sigma}^2\right),$$

where $U((0,j))_n/n$ is the local averaging estimator of the j-th covariance without bias correction and $\hat{\sigma}^2$ is an estimator of the IV; see Jacod et al. (2017) for more details. While Jacod et al. (2017) provide a finite sample bias correction when developing their local averaging estimators of noise covariances, they don't consider the feedback between, and unified by RV and LA based on 1,000 simulations. In the top panel we plot the estimators without finite sample bias correction and we plot the estimators with finite sample bias correction in the bottom panel, in which we use the true IV instead of any approximation/estimator to make the bias correction. We will analyze the case in which IV is estimated in the next subsection.

We observe that both estimators (RV and LA) perform poorly without finite sample bias correction. In particular, the noise autocorrelations estimated by the LA estimators decay slowly and hover above 0 up to 20 lags, from which we might conclude that the noise exhibits strong and long-memory dependence, while the underlying noise is, in fact, only weakly dependent. However, both estimators perform well after the finite sample bias correction. In Figure 3, we also plot the 95% simulated confidence intervals of the two bias corrected estimators. In terms of mean squared errors, both estimators, after bias correction, yield accurate estimates.

Figures 2-3 reveal that the finite sample bias correction is crucial to obtain reliable estimates of noise moments. The key ingredient of this correction, however, is (an estimate of) the IV. Yet, to obtain an estimate of the IV, we need to estimate the second moments of noise first — whence the feedback loop of bias corrections. This is why we introduced our multi-step estimators, which allow successive bias corrections in estimates for both the IV and noise autocorrelations.

6.3 Multi-step estimators of IV

In this subsection, we examine the performance of different estimators of the IV. We compare the estimator $\widehat{\mathrm{IV}}_n$ in (22) which is generated by the pre-averaging method using non-overlapping intervals, with the estimator $\widetilde{\mathrm{IV}}_n$ defined in (25) using overlapping intervals. We then assess the improvement in accuracy from our unified treatment of asymptotic and finite sample biases that can be achieved by using the K-step estimators $\widehat{\mathrm{IV}}_n^{(K)}$ and $\widetilde{\mathrm{IV}}_n^{(K)}$ introduced in (40). We also compare $\widehat{\mathrm{IV}}_n^{(1)}$ and $\widetilde{\mathrm{IV}}_n^{(1)}$ to $\widehat{\mathrm{IV}}_n^{(2)}$ and $\widetilde{\mathrm{IV}}_n^{(2)}$ to assess the gained accuracy by dropping the possibly misspecified assumption of independent noise.

In Table 1, we report the centered means of our estimators and the standard deviations (between parentheses), based on 1,000 simulations.[14] Throughout this subsection, the tuning parameter j_n is fixed at 20, we take $\ell_n = 10$ and $\theta = 0.4$, and use the triangular kernel. When comparing the estimators $\widehat{\mathrm{IV}}_n$ and $\widehat{\mathrm{IV}}_n^{(2)}$ in the first and the third rows of Table 1, we observe the important advantage of our multi-step estimators over the pre-averaging method that ignores the finite sample bias, since our estimators yield strongly improved accuracy. Furthermore, a comparison to the results for $\widehat{\mathrm{IV}}_n^{(1)}$ and $\widehat{\mathrm{IV}}_n^{(2)}$ in the second and third rows leads to the striking conclusion that ignoring the finite sample bias yields even

treatment of, asymptotic and finite sample biases, which is a key interest in this paper.

[14] The numbers are multiplied by 10^5.

more inaccuracy than ignoring the dependence in noise. This shows that one should be cautious when applying estimators without appropriate bias corrections even with data on a 1-sec time scale (i.e., 23,400 observations in a day of 6.5 trading hours). The "cost" of applying our multi-step estimators $\widehat{\mathrm{IV}}_n^{(K)}$ is the slightly larger standard deviations they induce. This increased uncertainty is introduced by correcting the "interlocked" bias. However, the reduction in bias strictly dominates the slight increase in standard deviations when noise is dependent. Therefore, the multi-step estimators have smaller mean-squared errors than their counterparts in the first two rows of Table 1. These standard deviations can be reduced by the use of overlapping intervals, as can be observed when we compare the standard deviations of $\widehat{\mathrm{IV}}_n^{(K)}$ with those of $\widetilde{\mathrm{IV}}_n^{(K)}$ (i.e., the first four rows in Table 1 and the next four rows). Although the centered means of the estimators become slightly worse when we adopt overlapping pre-averaging intervals, the significant reduction in the standard deviations implies a better overall performance under a mean-squared error criterion.

The estimator $\widetilde{\mathrm{IV}}_n^{\mathrm{JLZ}}$ recently proposed in Jacod et al. (2019), which corrects the asymptotic bias of pre-averaging estimators by local averaging but does not include a unified treatment of asymptotic and finite sample biases, performs better than the estimators $\widehat{\mathrm{IV}}_n$ and $\widetilde{\mathrm{IV}}_n$, but worse than all estimators with finite sample bias corrections. The method proposed in Da and Xiu (2019) generates an estimator $\widehat{\mathrm{IV}}_n^{\mathrm{QMLE}}$ which outperforms our method when the autocorrelation in the noise is small, but its performance deteriorates when the noise autocorrelation parameter ι is closer to -1 or 1.

In Table 2, we replicate the results of Table 1 but now with a higher data frequency, which corresponds to sampling every 0.2 seconds (i.e., 117,000 observations in a day of 6.5 trading hours). We observe that, with such very high-frequency data, the multi-step estimators still perform much better than their counterparts in rows 1 and 2, and 5 and 6, of Table 2 — with much smaller biases and only slightly larger standard deviations. Indeed, both the errors caused by ignoring the finite sample bias and the inconsistencies caused by a potential misspecification of the dependence structure in noise when using the first-step estimators remain clearly visible. The biases in the estimates typically reduce further when we replace $\widehat{\mathrm{IV}}_n^{(2)}$ by $\widehat{\mathrm{IV}}_n^{(3)}$, but not in all cases where $\widetilde{\mathrm{IV}}_n^{(2)}$ is replaced by $\widetilde{\mathrm{IV}}_n^{(3)}$. We also observe that increasing K in our multi-step estimators $\widehat{\mathrm{IV}}_n^{(K)}$ and $\widetilde{\mathrm{IV}}_n^{(K)}$ gives only a slight increase in the estimators' standard deviations. As before, the standard deviations of $\widetilde{\mathrm{IV}}_n^{(2)}$ and $\widetilde{\mathrm{IV}}_n^{(3)}$ are substantially smaller than for $\widehat{\mathrm{IV}}_n^{(2)}$ and $\widehat{\mathrm{IV}}_n^{(3)}$, and for $\widetilde{\mathrm{IV}}_n^{\mathrm{JLZ}}$ and $\widehat{\mathrm{IV}}_n^{\mathrm{QMLE}}$. In terms of CPU, the QMLE-estimator is relatively more time-consuming to compute. Indeed, in the setting of Table 2, 0.1% of the total computing time was spent on our four estimators based on non-overlapping intervals; 3.1% was spent on our four estimators based on overlapping intervals; 7.2% was spent on $\widetilde{\mathrm{IV}}_n^{\mathrm{JLZ}}$; and 89.6% was spent on $\widehat{\mathrm{IV}}_n^{\mathrm{QMLE}}$.

To numerically "verify" the central limit theorem established in Theorem 5.1, we plot the quantiles of the normalized estimators $\Delta_n^{-\frac{1}{4}}\left(\widehat{\mathrm{IV}}_n^{(2)}-\mathrm{IV}\right)/\sqrt{\widehat{\Sigma}_{\mathrm{IV}_n}^{(2)}}$ and $\Delta_n^{-\frac{1}{4}}\left(\widetilde{\mathrm{IV}}_n^{(2)}-\mathrm{IV}\right)/\sqrt{\widetilde{\Sigma}_{\mathrm{IV}_n}^{(2)}}$ against standard normal quantiles in Figure 4. We observe that the established limit distributions are clearly verified.

Remark 6.1 (Dependence between X and U)**.** *The theoretical results in this paper assume independence between X and U. In practice, the efficient price and the microstructure noise processes may be correlated. In Appendix B, we provide additional Monte Carlo simulation results that assess the effects of price discreteness and correlation between X and U. (Price discreteness renders dependence between X and U.) Our results show that the presence of minimal ticks has relatively little impact on the estimation of the moments of noise and the IV. Furthermore, our results show that in the situation when X and U are correlated our multi-step estimators still appear to be performing well.*

7 Empirical Study

7.1 Data description

We analyze the NYSE TAQ transaction prices of Citigroup (trading symbol: C) over the month January 2011. We discard all transactions before 9:30 and after 16:00. We retain a total of 4,933,059 transactions over 20 trading days, thus on average 10.5 observations per second. The estimation is first performed on the full sample, and then on subsamples obtained by different sampling schemes. We demonstrate how the sampling methods affect the properties of the noise, and thus affect the estimation of the IV. We employ pre-averaging on overlapping intervals, and use the triangular kernel. Throughout this section, the tuning parameter of the RV estimator is fixed at $j_n = 30$ and θ is selected according to the optimal rule (28).

7.2 Estimating the second moments of noise

We estimate the j-th autocovariance and autocorrelation of microstructure noise with $j = 0, 1, \ldots, 30$ by three estimators: our realized volatility (RV) estimators in (7) and (8), the local averaging (LA) estimators proposed by Jacod et al. (2017), and the bias corrected realized volatility (BCRV) estimators in (37) and (38). We perform the estimation over each trading day and end up with 20 estimates (of the 30 lags of autocovariances or autocorrelations) for each estimator. In Figure 5 we plot the average of the 20 estimates (over the month) as well as the approximated confidence intervals that are two sample standard deviations away from the mean.

We observe that the three estimators yield quite close estimates by virtue of the high data frequency. Noise in this sample tends to be positively autocorrelated — with the BCRV estimators yielding the fastest decay. Empirically this positive autocorrelation is consistent with the finding that the arrivals of buy and sell orders are positively autocorrelated; see Hasbrouck and Ho (1987). This corresponds to the trading practice that informed traders split their orders over (a short period of) time and trade on one side of the market, rendering continuation in their orders.

We emphasize that the finite sample bias can be much more pronounced than what we observe in Figure 5, even if we conduct estimation on a full transaction data sample. Indeed, in Appendix C, we analyze a sample of transaction prices of General Electric (GE) and show that, when the data frequency is very high, the finite sample bias correction is particularly essential when the noise-to-signal ratio is very small (recall Remark 3.3).

7.3 Estimating the IV

Turning to the estimation of the IV, we study four estimators of the pre-averaging class: $\widetilde{\mathrm{IV}}_n$, $\widetilde{\mathrm{IV}}_n^{(1)}$, $\widetilde{\mathrm{IV}}_n^{(2)}$, and $\widetilde{\mathrm{IV}}_n^{\mathrm{JLZ}}$. In the top panel of Figure 6, we plot the four estimators of the IV for each trading day. We note that the three estimators $\widetilde{\mathrm{IV}}_n$, $\widetilde{\mathrm{IV}}_n^{\mathrm{JLZ}}$, and $\widetilde{\mathrm{IV}}_n^{(2)}$ yield quite close results. This is expected, as the three methods, RV, LA, and BCRV, provide close estimates of the second moments of noise. However, the estimator $\widetilde{\mathrm{IV}}_n^{(1)}$, which ignores the dependence in noise, yields very different estimates, and the differences are persistent — $\widetilde{\mathrm{IV}}_n^{(1)}$ yields higher estimates over each trading day. Moreover, the differences are statistically significant by virtue of Theorem 5.1 — all the 20 estimates fall outside of the 95% confidence intervals, as the bottom panel of Figure 6 reveals.

7.4 Decaying rate of autocorrelation

Figure 5 shows that the positive autocorrelations of noise drop to zero rapidly. To assess the rate of decay, we perform a logarithmic transformation of the autocorrelations estimated by BCRV.[15] In Figure 7, we plot the logarithmic autocorrelations for each trading day (top panel) and the mean logarithmic autocorrelations over all days (bottom panel). The plots indicate that the logarithmic autocorrelation is approximately a linear function of the number of lags, i.e., the autocorrelation function is decaying at an exponential rate.[16]

7.5 Robustness check — estimation under other sampling schemes

It is interesting to analyze how our estimators perform when the data is sampled at different time scales. In this section, we consider two alternative (sub)sampling schemes: regular time sampling and tick time sampling (recall Remark 2.1 for details on the sampling schemes).

[15] We restrict attention to the lags up to $j = 10$. The logarithmic autocorrelations at higher lags are very volatile since the autocorrelations are close to zero.

[16] The autocorrelation decay rate would be slower without unified treatment of the bias corrections, which may explain the relatively strong polynomial dependence in noise found in Jacod et al. (2017) and questioned by these authors themselves.

7.5.1 Regular time sampling

The prices in this sample are recorded on a 1-second time scale. If there were multiple prices in a second, we select the first one; and we do not record a price if there is no transaction in a second. We end up with 21,691 observations on average per trading day. Figure 8 is analogous to Figure 5. The three estimators, RV, LA, and BCRV, now produce very different patterns. Both the RV and LA estimators suggest that the noise is strongly autocorrelated in this subsample, even stronger than in the original full sample. This would be counterintuitive since we eliminate more than 90% of the full sample in a fairly random way — the elimination should have weakened the serial dependence of noise in the remaining sample. However, the estimates by BCRV reveal that in fact the noise is approximately uncorrelated — it is the finite sample bias that makes the autocorrelations of noise seem strong and persistent if not taken into account.

If the noise is close to being independent, then $\widetilde{\mathrm{IV}}_n^{(1)}$, which assumes i.i.d. noise, would be a sound estimator of the IV. An alternative estimator, e.g., $\widetilde{\mathrm{IV}}_n^{(2)}, \widetilde{\mathrm{IV}}_n$, or $\widetilde{\mathrm{IV}}_n^{\mathrm{JLZ}}$ would then be robust if it would deliver similar estimates. In the top panel of Figure 9, we observe that $\widetilde{\mathrm{IV}}_n^{(1)}$ and $\widetilde{\mathrm{IV}}_n^{(2)}$ yield virtually identical estimates. The other two estimators, $\widetilde{\mathrm{IV}}_n$ and $\widetilde{\mathrm{IV}}_n^{\mathrm{JLZ}}$ which don't apply finite sample bias corrections, however, yield even negative estimates. It is interesting to briefly elaborate on the performance of $\widetilde{\mathrm{IV}}_n$ and $\widetilde{\mathrm{IV}}_n^{\mathrm{JLZ}}$ in this scenario. Using the triangular kernel, with the selected tuning parameters $j_n = 30, \ell_n = 4, K_n = 7$ and a reasonable choice of $\theta = 0.2$, we have by (43) that $\mathrm{Err}_{\mathrm{JLZ}} = 103.69\%$, $\mathrm{Err}_{\mathrm{RV}} = 175.64\%$. Therefore, $\widetilde{\mathrm{IV}}_n^{\mathrm{JLZ}}$ and $\widetilde{\mathrm{IV}}_n$ are in fact estimating -3.69% and -75.64% of the IV, and this is in line with the estimates in the top panel of Figure 9. We conclude that Figures 6 and 9 jointly confirm the importance of our multi-step approach. Indeed, $\widetilde{\mathrm{IV}}_n^{(1)}$, which assumes i.i.d. noise, exhibits unreliable behavior in Figure 6, while $\widetilde{\mathrm{IV}}_n$, which doesn't apply finite sample bias corrections, shows unreliable behavior in Figure 9.

7.5.2 Tick time sampling

In a tick time sample, prices are collected with each price change, i.e., all zero returns are suppressed, see, e.g., Zhou (1996), Griffin and Oomen (2008), Aït-Sahalia et al. (2011), Kalnina (2011) and Da and Xiu (2019). For the Citigroup transaction data, 70% of the returns are zero. The corresponding average number of prices per second in our tick time sample is 3.2. Figure 10 shows that the microstructure noise has a different dependence pattern in the tick time sample — its autocorrelation function is alternating. Masked by alternating noise, the observed returns at tick time have a similar pattern; see Aït-Sahalia et al. (2011) and Griffin and Oomen (2008). This dependence structure of noise is perceived to be due to the discreteness of price changes, irrespective of the distributional features of noise in the original transactions or quotes data.

Interestingly, the bottom panel of Figure 9 shows that the two estimators of the IV, $\widetilde{\mathrm{IV}}_n^{(1)}$ and $\widetilde{\mathrm{IV}}_n^{(2)}$, remain close. It is not immediate why $\widetilde{\mathrm{IV}}_n^{(1)}$ and $\widetilde{\mathrm{IV}}_n^{(2)}$ deliver almost identical estimates, given the fact that the dependence of noise in this tick time sample is drastically different from i.i.d. noise. However, a clue is provided by the observation that negatively autocorrelated noise has less impact on the estimation of the IV, as the high-order alternating autocovariances partially cancel out, and thus contribute less to the asymptotic bias σ_U^2.[17] $\widetilde{\mathrm{IV}}_n$ and $\widetilde{\mathrm{IV}}_n^{\mathrm{JLZ}}$ are still significantly underestimating the IV due to the finite sample bias.

7.6 Economic interpretation and empirical implication

The dependence structure of microstructure noise depends on the sampling frequency and scheme. In an original transaction data sample, noise is likely to be positively autocorrelated as a result of various trading practices that entail continuation in order flows. The dependence of noise can be reduced by sampling sparsely, say, every few (or more) seconds as we show in Section 7.5.1; noise is close to independent in such sparse subsamples. If, however, we remove all zero returns, thus sample in tick time, noise typically exhibits an alternating autocorrelogram.

Microstructure theories can provide some intuitive economic interpretations of the dynamic properties of microstructure noise recovered in this paper. The positive autocorrelation function displayed in Figure 5 is consistent with the findings in Hasbrouck and Ho (1987), Choi et al. (1988) and Huang and Stoll (1997) that explicitly model the probability of order reversal π (or order continuation by $1-\pi$),[18] so that the deviation of transaction prices from fundamentals becomes an AR(1) process. Fitting the autocorrelation function recovered by BCRV in Figure 5 to that of an AR(1) model, we obtain an estimate of the AR(1) coefficient equal to $\hat{\iota} = 0.73$ and the probability of order continuation is $1-\hat{\pi} = (1+\hat{\iota})/2 = 0.87$. That is, the estimated probability that a buy (or sell) order follows another buy (or sell) order is 0.87. In view of the extensive empirical results in Huang and Stoll (1997) (see Table 5 therein), this is a reasonable estimate.

One possible interpretation of the positively autocorrelated order flows is that a large *order* is often executed as a series of smaller *trades* to reduce the price impact, or conducted against multiple *trades* from stale limit orders. However, such positive autocorrelation contradicts the prediction of inventory

[17] For a tractable analysis, one may consider AR(1) noise processes. Let $\iota \in (0,1)$ be the absolute value of the AR(1) coefficient. When the noise is positively autocorrelated, the asymptotic bias σ_U^2 corrected by $\widetilde{\mathrm{IV}}_n^{(1)}$ and $\widetilde{\mathrm{IV}}_n^{(2)}$ is $(1-\iota)\gamma(0)$ and $\frac{1+\iota}{1-\iota}\gamma(0)$, respectively; when the noise is negatively autocorrelated, it is $(1+\iota)\gamma(0)$ and $\frac{1-\iota}{1+\iota}\gamma(0)$. Consider $\iota = 0.8$. Then, $(1-\iota)\gamma(0) = 0.2\gamma(0)$ and $\frac{1+\iota}{1-\iota}\gamma(0) = 9\gamma(0)$ while $(1+\iota)\gamma(0) = 1.8\gamma(0)$ and $\frac{1-\iota}{1+\iota}\gamma(0) = \frac{1}{9}\gamma(0)$. Therefore, the difference in the asymptotic bias is smaller when the noise is negatively autocorrelated; consequently, the IV estimates by $\widetilde{\mathrm{IV}}_n^{(1)}$ and $\widetilde{\mathrm{IV}}_n^{(2)}$ are close, see also Tables 1 and 2 in our simulation study.

[18] It is the probability that a buy (sell) order follows another sell (buy) order.

models, in which market makers induce negatively autocorrelated order flows to equilibrate inventories; see Ho and Stoll (1981). Consequently, according to inventory models the probability of order reversal would be $\pi > 0.5$. One remedy, suggested by Huang and Stoll (1997), is to collapse multiple *trades* at the same price into one *order*, which is exactly the tick time sampling scheme considered in Section 7.5.2. Exploiting the estimates by BCRV presented in Figure 10, we obtain an estimate of the probability of order reversal equal to $\hat{\pi} = 0.84$, which is very close to the average probability 0.87 in Huang and Stoll (1997). We emphasize that we recover these probabilities without any prior knowledge or estimates of the order flows.

The dependence structure of microstructure noise affects the estimation of the IV. Popular de-noise methods that assume i.i.d. noise work reasonably well with relatively sparse regular time samples or tick time samples. However, this discards a substantial amount of the original transaction data.[19] Instead, we can directly estimate the IV from the original data using our multi-step estimators that explicitly take the potential dependence in noise into account.

In our empirical study, we have also illustrated that bias corrections play an essential role in recovering the statistical properties of noise and in estimating the IV. Our multi-step estimators are specifically designed to conduct such bias corrections.

8 Conclusion

In high-frequency financial data the efficient price is contaminated by microstructure noise, which is usually assumed to be independently and identically distributed. This simple distributional assumption is challenged by both microeconomic financial models and various empirical facts. In this paper, we deviate from the i.i.d. assumption by allowing noise to be dependent in a general setting. We then develop econometric tools to recover the dynamic properties of microstructure noise and design improved approaches for the estimation of the integrated volatility.

This paper makes four contributions. First, it develops nonparametric estimators of the second moments of microstructure noise in a general setting. Second, it provides robust estimators of the integrated volatility, without assuming serially independent noise. Third, it reveals the importance of both asymptotic and finite sample bias analysis and develops simple and readily implementable multi-step estimators. Empirically, it characterizes the dependence structures of noise at several time scales and provides intuitive economic interpretations; it also investigates the impact of the dynamic properties of microstructure noise on integrated volatility estimation.

[19] To obtain the Citigroup tick time sample and the 1-second regular time sample, we delete roughly 70% and 90% of the original transaction data, respectively.

This paper thus introduces a robust and accurate method to effectively separate the two components of high-frequency financial data — the efficient price and microstructure noise. The robustness lies in its flexibility to accommodate rich dependence structures of microstructure noise motivated by various economic models and trading practices, whereas the accuracy is achieved by the finite sample refinement. As a result, we discover dynamic properties of microstructure noise consistent with microstructure theory and obtain accurate volatility estimators.

Acknowledgements

We are very grateful to the Editor, the Associate Editor and two anonymous referees for their helpful comments and suggestions that have significantly improved this paper. We are also grateful to Federico Bandi, Peter Boswijk, Peter Reinhard Hansen, Siem Jan Koopman, Oliver Linton, and Xiye Yang for their comments and discussions on earlier versions of this paper. This research was funded in part by the Netherlands Organization for Scientific Research under grant NWO VIDI 2009 (Laeven).

References

Aït-Sahalia, Y. and J. Jacod (2014): *High-frequency Financial Econometrics*, Princeton University Press.

Aït-Sahalia, Y., P. A. Mykland, and L. Zhang (2005): "How often to sample a continuous-time process in the presence of market microstructure noise," *Review of Financial Studies*, 18, 351–416.

——— (2011): "Ultra high frequency volatility estimation with dependent microstructure noise," *Journal of Econometrics*, 160, 160–175.

Bandi, F. M. and J. R. Russell (2006): "Separating microstructure noise from volatility," *Journal of Financial Economics*, 79, 655–692.

——— (2008): "Microstructure noise, realized variance, and optimal sampling," *Review of Economic Studies*, 75, 339–369.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2008): "Designing realized kernels to measure the ex post variation of equity prices in the presence of noise," *Econometrica*, 76, 1481–1536.

Black, F. (1986): "Noise," *Journal of Finance*, 41, 529–543.

Bradley, R. C. (2007): *Introduction to Strong Mixing Conditions*, Kendrick Press.

Chaker, S. (2017): “On high frequency estimation of the frictionless price: The use of observed liquidity variables,” *Journal of Econometrics*, 201, 127–143.

Choi, J. Y., D. Salandro, and K. Shastri (1988): “On the estimation of bid-ask spreads: Theory and evidence,” *Journal of Financial and Quantitative Analysis*, 23, 219–230.

Christensen, K., R. C. Oomen, and M. Podolskij (2014): “Fact or friction: Jumps at ultra high frequency,” *Journal of Financial Economics*, 114, 576–599.

Clinet, S. and Y. Potiron (2017): “Estimation for high-frequency data under parametric market microstructure noise,” Tech. rep.

——— (2018): “Efficient asymptotic variance reduction when estimating volatility in high frequency data,” *Journal of Econometrics*, 206, 103–142.

——— (2019): “Testing if the market microstructure noise is fully explained by the informational content of some variables from the limit order book,” *Journal of Econometrics*.

Da, R. and D. Xiu (2019): “When Moving-Average Models Meet High-Frequency Data: Uniform Inference on Volatility,” Tech. rep.

Duffie, D. (2010): *Dynamic Asset Pricing Theory*, Princeton University Press.

Garman, M. B. (1976): “Market microstructure,” *Journal of Financial Economics*, 3, 257–275.

Glosten, L. R. and P. R. Milgrom (1985): “Bid, ask and transaction prices in a specialist market with heterogeneously informed traders,” *Journal of Financial Economics*, 14, 71–100.

Gloter, A. and J. Jacod (2001a): “Diffusions with measurement errors. I. Local asymptotic normality,” *ESAIM: Probability and Statistics*, 5, 225–242.

——— (2001b): “Diffusions with measurement errors. II. Optimal estimators,” *ESAIM: Probability and Statistics*, 5, 243–260.

Griffin, J. E. and R. C. Oomen (2008): “Sampling returns for realized variance calculations: tick time or transaction time?” *Econometric Reviews*, 27, 230–253.

Gross-KlussMann, A. and N. Hautsch (2013): “Predicting bid–ask spreads using long-memory autoregressive conditional Poisson models,” *Journal of Forecasting*, 32, 724–742.

Hansen, P. R., J. Large, and A. Lunde (2008): “Moving average-based estimators of integrated variance,” *Econometric Reviews*, 27, 79–111.

Hansen, P. R. and A. Lunde (2006): "Realized variance and market microstructure noise," *Journal of Business & Economic Statistics*, 24, 127–161.

Harris, L. (1990): "Estimation of stock price variances and serial covariances from discrete observations," *Journal of Financial and Quantitative Analysis*, 25, 291–306.

Hasbrouck, J. (2007): *Empirical Market Microstructure: The Institutions, Economics, and Econometrics of Securities Trading*, Oxford University Press.

Hasbrouck, J. and T. S. Ho (1987): "Order arrival, quote behavior, and the return-generating process," *Journal of Finance*, 42, 1035–1048.

Hautsch, N. and M. Podolskij (2013): "Preaveraging-based estimation of quadratic variation in the presence of noise and jumps: Theory, implementation, and empirical evidence," *Journal of Business & Economic Statistics*, 31, 165–183.

Hendershott, T., C. M. Jones, and A. J. Menkveld (2013): "Implementation shortfall with transitory price effects," in *High-Frequency Trading: New Realities for Trades, Markets and Regulators*, ed. by D. Easley, M. L. de Prado, and M. O'Hara, London: Risk Books.

Ho, T. and H. R. Stoll (1981): "Optimal dealer pricing under transactions and return uncertainty," *Journal of Financial Economics*, 9, 47–73.

Huang, R. D. and H. R. Stoll (1997): "The components of the bid-ask spread: A general approach," *Review of Financial Studies*, 10, 995–1034.

Jacod, J., Y. Li, P. A. Mykland, M. Podolskij, and M. Vetter (2009): "Microstructure noise in the continuous case: The pre-averaging approach," *Stochastic Processes and Their Applications*, 119, 2249–2276.

Jacod, J., Y. Li, and X. Zheng (2017): "Statistical properties of microstructure noise," *Econometrica*, 85, 1133–1174.

——— (2019): "Estimating the integrated volatility with tick observations," *Journal of Econometrics*, 208, 80–100.

Jacod, J. and P. A. Mykland (2015): "Microstructure noise in the continuous case: Approximate efficiency of the adaptive pre-averaging method," *Stochastic Processes and their Applications*, 125, 2910–2936.

Jacod, J., M. Podolskij, and M. Vetter (2010): "Limit theorems for moving averages of discretized processes plus noise," *Annals of Statistics*, 38, 1478–1545.

Jacod, J. and A. N. Shiryaev (2003): *Limit Theorems for Stochastic Processes*, vol. 288, Springer-Verlag Berlin.

Kalnina, I. (2011): "Subsampling high frequency data," *Journal of Econometrics*, 161, 262–283.

Lee, S. S. and P. A. Mykland (2012): "Jumps in equilibrium prices and market microstructure noise," *Journal of Econometrics*, 168, 396–406.

Li, Y., S. Xie, and X. Zheng (2016): "Efficient estimation of integrated volatility incorporating trading information," *Journal of Econometrics*, 195, 33–50.

Li, Z. M., R. J. A. Laeven, and M. H. Vellekoop (2019): "Supplementary material to "Dependent microstructure noise and integrated volatility estimation from high-frequency data," Tech. rep.

Mancini, C. (2001): "Disentangling the jumps of the diffusion in a geometric jumping Brownian motion," *Giornale dell'Istituto Italiano degli Attuari*, 64, 19–47.

Mokkadem, A. (1988): "Mixing properties of ARMA processes," *Stochastic Processes and Their Applications*, 29, 309–315.

Mykland, P. A. and L. Zhang (2012): "The econometrics of high frequency data," in *Statistical Methods for Stochastic Differential Equations*, ed. by M. Kessler, A. Lindner, and M. Sörensen, New York: Chapman and Hall/CRC Press, chap. 2, 109–190.

Podolskij, M. and M. Vetter (2009a): "Bipower-type estimation in a noisy diffusion setting," *Stochastic Processes and Their Applications*, 119, 2803–2831.

——— (2009b): "Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps," *Bernoulli*, 15, 634–658.

Roll, R. (1984): "A simple implicit measure of the effective bid-ask spread in an efficient market," *Journal of Finance*, 39, 1127–1139.

Stoll, H. R. (1989): "Inferring the components of the bid-ask spread: theory and empirical tests," *Journal of Finance*, 44, 115–134.

Tsay, R. S. (2005): *Analysis of Financial Time Series*, vol. 543, John Wiley & Sons.

Xiu, D. (2010): "Quasi-maximum likelihood estimation of volatility with high frequency data," *Journal of Econometrics*, 159, 235–250.

Zhang, L. (2006): "Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach," *Bernoulli*, 12, 1019–1043.

Zhang, L., P. A. Mykland, and Y. Aït-Sahalia (2005): "A tale of two time scales: Determining integrated volatitility with noisy high-frequency data," *Journal of the American Statistical Association*, 100, 1394–1411.

Zhou, B. (1996): "High-frequency data and volatility in foreign-exchange rates," *Journal of Business & Economic Statistics*, 14, 45–52.

Tables and Figures

ι	-0.7	-0.3	0	0.3	0.7
$\widehat{\mathrm{IV}}_n$	-22.37 (14.15)	-22.36 (14.17)	-22.36 (14.18)	-22.40 (14.20)	-22.87 (14.27)
$\widehat{\mathrm{IV}}_n^{(1)}$	-1.71 (4.19)	-0.97 (4.21)	-0.23 (4.24)	0.85 (4.29)	4.33 (4.47)
$\widehat{\mathrm{IV}}_n^{(2)}$	-0.94 (5.58)	-0.55 (5.61)	-0.19 (5.65)	0.31 (5.72)	1.57 (5.98)
$\widehat{\mathrm{IV}}_n^{(3)}$	-0.55 (6.32)	-0.35 (6.35)	-0.17 (6.40)	0.04 (6.48)	0.19 (6.79)
$\widetilde{\mathrm{IV}}_n$	-22.66 (13.94)	-22.66 (13.96)	-22.68 (13.96)	-22.74 (13.97)	-23.30 (13.99)
$\widetilde{\mathrm{IV}}_n^{(1)}$	-2.00 (3.07)	-1.27 (3.08)	-0.55 (3.09)	0.51 (3.10)	3.90 (3.18)
$\widetilde{\mathrm{IV}}_n^{(2)}$	-1.37 (3.60)	-1.01 (3.60)	-0.67 (3.60)	-0.20 (3.61)	0.93 (3.69)
$\widetilde{\mathrm{IV}}_n^{(3)}$	-1.06 (3.89)	-0.88 (3.89)	-0.73 (3.90)	-0.56 (3.91)	-0.55 (3.99)
$\widetilde{\mathrm{IV}}_n^{\mathrm{JLZ}}$	-11.74 (7.63)	-11.65 (7.63)	-11.65 (7.64)	-11.65 (7.65)	-11.19 (7.68)
$\widehat{\mathrm{IV}}_n^{\mathrm{QMLE}}$	0.83 (10.13)	-0.19 (3.16)	-0.18 (3.43)	0.04 (3.52)	1.08 (4.35)

Table 1: Estimation of the IV. We take $\Delta = 1$ sec and the number of observations is $n = 23{,}400$. We report the estimation results of three groups of IV estimators: our pre-averaging estimator and its multi-step versions based on non-overlapping intervals $\widehat{\mathrm{IV}}_n, \widehat{\mathrm{IV}}_n^{(1)}, \widehat{\mathrm{IV}}_n^{(2)}$ and $\widehat{\mathrm{IV}}_n^{(3)}$; our pre-averaging estimator and its multi-step versions based on overlapping intervals $\widetilde{\mathrm{IV}}_n, \widetilde{\mathrm{IV}}_n^{(1)}, \widetilde{\mathrm{IV}}_n^{(2)}$ and $\widetilde{\mathrm{IV}}_n^{(3)}$; the estimator $\widetilde{\mathrm{IV}}_n^{\mathrm{JLZ}}$ based on the pre-averaging method proposed in Jacod et al. (2019) and the estimator $\widehat{\mathrm{IV}}_n^{\mathrm{QMLE}}$ based on the QMLE method in Da and Xiu (2019). The numbers represent the centered mean estimates based on 1,000 simulations with standard deviations between parentheses. All numbers in the table have been multiplied by 10^5. The tuning parameters for the first eight estimators are $j_n = 20$, $\ell_n = 10$ and $\theta = 0.4$, and we use the triangular kernel. For the estimator in Jacod et al. (2019) we used the choices suggested in that paper: $\bar{h}_n = 0.5/\sqrt{\Delta_n}$, $k_n = (\Delta_n)^{-1/5}$ and $k'_n = (\Delta_n)^{-1/8}$. In Da and Xiu (2019) the parameter q of the fitted MA(q) model was found by optimization over $q \in \{8, 9, 10\}$ only for each sample in order to save time, since test runs indicated that the optimal order was usually around $q = 9$.

ι	-0.7	-0.3	0	0.3	0.7
$\widehat{\mathrm{IV}}_n$	-4.49 (3.87)	-4.49 (3.88)	-4.49 (3.90)	-4.50 (3.93)	-4.62 (4.05)
$\widehat{\mathrm{IV}}_n^{(1)}$	-1.47 (2.83)	-0.72 (2.85)	0.02 (2.87)	1.13 (2.91)	4.98 (3.06)
$\widehat{\mathrm{IV}}_n^{(2)}$	-0.11 (3.07)	-0.03 (3.09)	0.04 (3.11)	0.13 (3.15)	0.40 (3.32)
$\widehat{\mathrm{IV}}_n^{(3)}$	0.02 (3.09)	0.04 (3.12)	0.04 (3.14)	0.03 (3.18)	-0.06 (3.35)
$\widetilde{\mathrm{IV}}_n$	-4.83 (3.48)	-4.83 (3.48)	-4.83 (3.49)	-4.85 (3.50)	-5.00 (3.55)
$\widetilde{\mathrm{IV}}_n^{(1)}$	-1.80 (2.13)	-1.06 (2.13)	-0.32 (2.14)	0.78 (2.15)	4.60 (2.20)
$\widetilde{\mathrm{IV}}_n^{(2)}$	-0.48 (2.28)	-0.41 (2.29)	-0.34 (2.29)	-0.25 (2.31)	-0.02 (2.37)
$\widetilde{\mathrm{IV}}_n^{(3)}$	-0.35 (2.30)	-0.34 (2.30)	-0.34 (2.31)	-0.36 (2.32)	-0.48 (2.39)
$\widetilde{\mathrm{IV}}_n^{\mathrm{JLZ}}$	-3.79 (3.11)	-3.68 (3.12)	-3.68 (3.12)	-3.68 (3.13)	-3.09 (3.17)
$\widehat{\mathrm{IV}}_n^{\mathrm{QMLE}}$	0.50 (3.61)	-0.69 (2.64)	-0.76 (3.16)	-0.80 (3.28)	0.28 (4.74)

Table 2: Estimation of the IV. We take $\Delta = 0.2$ sec and the number of observations is $n = 117{,}000$. We report the estimation results of three groups of IV estimators: our pre-averaging estimator and its multi-step versions based on non-overlapping intervals $\widehat{\mathrm{IV}}_n, \widehat{\mathrm{IV}}_n^{(1)}, \widehat{\mathrm{IV}}_n^{(2)}$ and $\widehat{\mathrm{IV}}_n^{(3)}$; our pre-averaging estimator and its multi-step versions based on overlapping intervals $\widetilde{\mathrm{IV}}_n, \widetilde{\mathrm{IV}}_n^{(1)}, \widetilde{\mathrm{IV}}_n^{(2)}$ and $\widetilde{\mathrm{IV}}_n^{(3)}$; the estimator $\widetilde{\mathrm{IV}}_n^{\mathrm{JLZ}}$ based on the pre-averaging method proposed in Jacod et al. (2019) and the estimator $\widehat{\mathrm{IV}}_n^{\mathrm{QMLE}}$ based on the QMLE method in Da and Xiu (2019). The numbers represent the centered mean estimates based on 1,000 simulations with standard deviations between parentheses. All numbers in the table have been multiplied by 10^5. The tuning parameters for the first eight estimators are $j_n = 20$, $\ell_n = 10$ and $\theta = 0.4$, and we use the triangular kernel. For the estimator in Jacod et al. (2019) we used the choices suggested in that paper: $\bar{h}_n = 0.5/\sqrt{\Delta_n}$, $k_n = (\Delta_n)^{-1/5}$ and $k_n' = (\Delta_n)^{-1/8}$. In Da and Xiu (2019) the parameter q of the fitted MA(q) model was found by optimization over $q \in \{8, 9, 10\}$ only for each sample in order to save time, since test runs indicated that the optimal order was usually around $q = 9$.

$$
\begin{tikzcd}
\textbf{Step 1:} & {} \arrow[r, "\text{Observed price } Y", "\text{RV}"'] & \widetilde{\gamma(0)}_n^{(1)} \arrow[r, "\text{Asymptotic bias}", "\text{PAV}"'] & \widetilde{\mathrm{IV}}_n^{(1)} \arrow[d, "\text{Finite sample bias}"', "\text{RV}"] \\
\textbf{Step 2:} & & \widetilde{\mathrm{IV}}_n^{(2)} \arrow[d, "\text{Finite sample bias}"', "\text{RV}"] & \widetilde{\gamma(0)}_n^{(2)}, \widetilde{\gamma(j)}_n^{(2)} \arrow[l, "\text{Asymptotic bias}"', "\text{PAV}"] \\
\textbf{Step 3:} & & \widetilde{\gamma(0)}_n^{(3)}, \widetilde{\gamma(j)}_n^{(3)} \arrow[r, "\text{Asymptotic bias}", "\text{PAV}"'] & \widetilde{\mathrm{IV}}_n^{(3)} \\
\vdots & & \vdots & \vdots \\
\textbf{Step } \boldsymbol{K}\textbf{:} & & \widetilde{\gamma(0)}_n^{(K)}, \widetilde{\gamma(j)}_n^{(K)} & \widetilde{\mathrm{IV}}_n^{(K)}
\end{tikzcd}
$$

Figure 1: Illustration of the construction of the multi-step estimators. In the first step, we use realized volatility (RV) to obtain an estimator of the variance of (possibly misspecified) i.i.d. noise, $\widetilde{\gamma(0)}_n^{(1)}$. Next, this estimator is used to correct the asymptotic bias of the pre-averaging estimator (PAV) to derive the first-step estimator of the IV, $\widetilde{\mathrm{IV}}_n^{(1)}$. In the second step, we use $\widetilde{\mathrm{IV}}_n^{(1)}$ to obtain finite sample bias corrected estimators of the variance and covariances of noise, $\widetilde{\gamma(0)}_n^{(2)}$ and $\widetilde{\gamma(j)}_n^{(2)}$, which are then used to remove the asymptotic bias in PAV, leading to the second-step IV estimator, $\widetilde{\mathrm{IV}}_n^{(2)}$. Iterating this procedure will lead to K-step estimators $\widetilde{\gamma(0)}_n^{(K)}$, $\widetilde{\gamma(j)}_n^{(K)}$, $\widetilde{\mathrm{IV}}_n^{(K)}$.

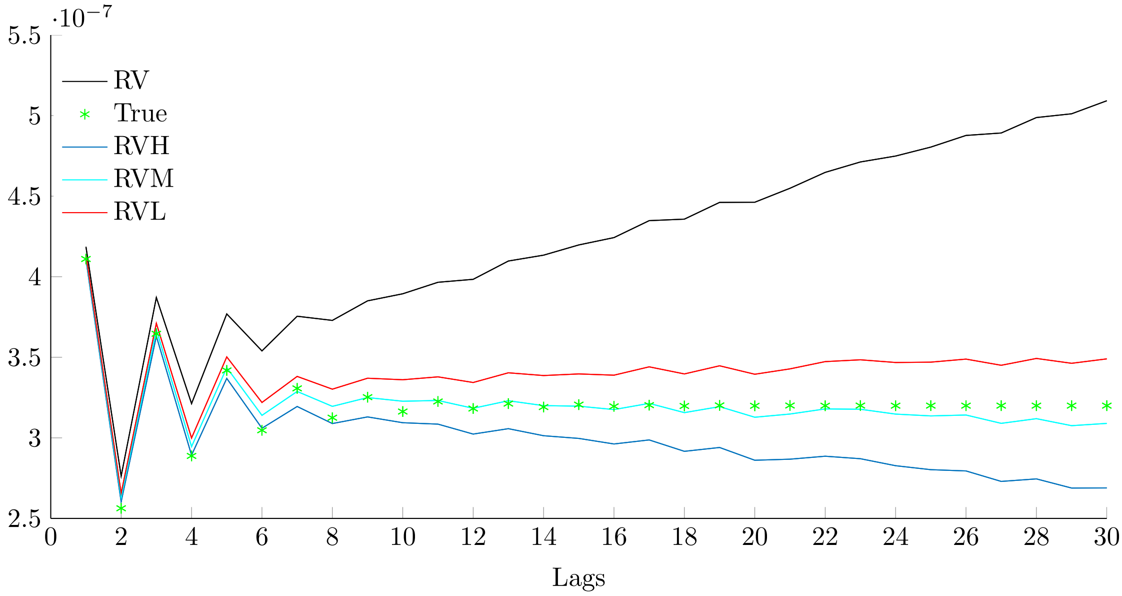

Figure 2: Realized volatility estimators against the number of lags j, based on a single simulated sample, without and with finite sample bias correction, cf. (6) and (10). Here, RV: $\widehat{\langle Y,Y\rangle}(j)_n$; RVL: $\widehat{\langle Y,Y\rangle}(j)_n - \frac{0.8j\mathrm{IV}}{2(n-j+1)}$; RVM: $\widehat{\langle Y,Y\rangle}(j)_n - \frac{j\mathrm{IV}}{2(n-j+1)}$; and RVH: $\widehat{\langle Y,Y\rangle}(j)_n - \frac{1.2j\mathrm{IV}}{2(n-j+1)}$. We take $\Delta = 1$ sec, the number of observations is 23,400, and $\iota = -0.7$. The designation "True" corresponds to the stochastic limit $\gamma(0) - \gamma(j)$.

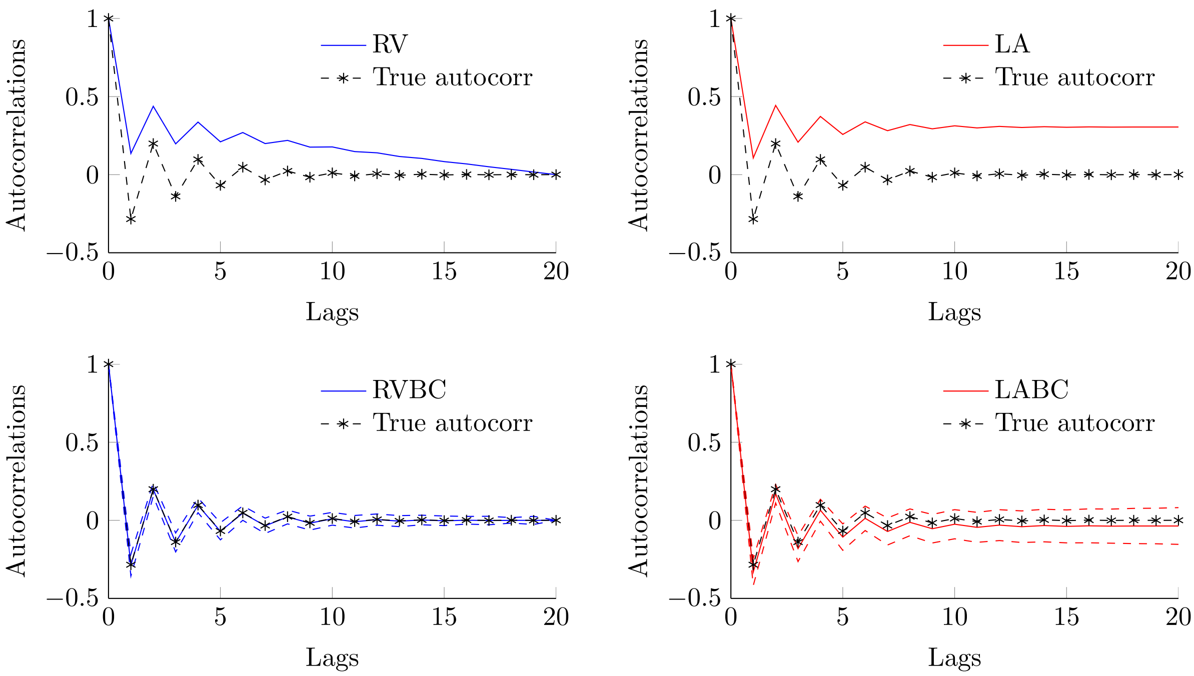

Figure 3: Realized volatility (RV) and local averaging (LA) estimators of the autocorrelations of noise against the number of lags j, averaged over 1,000 simulated samples. Top panel: RV and LA estimators without finite sample bias corrections. Bottom panel: RV and LA estimators with finite sample bias corrections (RVBC, LABC). The dashed lines are the 95% simulated confidence intervals. We take $\Delta = 1$ sec, the number of observations is 23,400, and $\iota = -0.7$. The tuning parameters of the RV and LA estimators are $j_n = 20$ and $K_n = 6$, respectively.

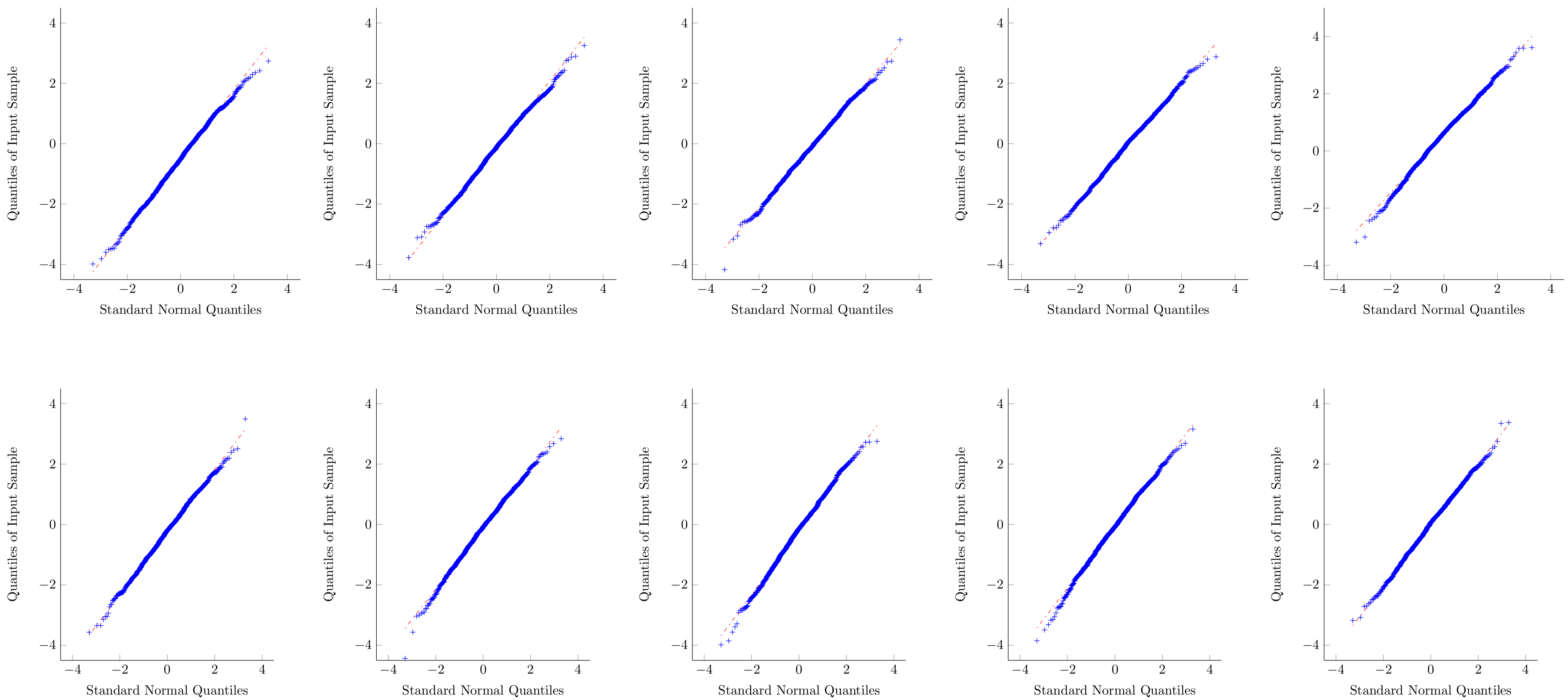

Figure 4: Standard normal QQ-plots of the second-step IV estimators. Top panel: $\Delta_n^{-\frac{1}{4}}\left(\widehat{\mathrm{IV}}_n^{(2)}-\mathrm{IV}\right)/\sqrt{\widehat{\Sigma}^{(2)}_{\mathrm{IV}_n}}$. Bottom panel: $\Delta_n^{-\frac{1}{4}}\left(\widetilde{\mathrm{IV}}_n^{(2)}-\mathrm{IV}\right)/\sqrt{\widetilde{\Sigma}^{(2)}_{\mathrm{IV}_n}}$. The AR(1) coefficient ι of the noise process, from left to right, is $-0.7, -0.3, 0, 0.3$, and 0.7. The number of simulations is 1,000, the data frequency is $\Delta = 0.1$ sec, and the number of observations is 234,000. The tuning parameter of the RV estimator is $j_n = 20$, and $\ell_n = 10$. The tuning parameter θ equals 1/3.

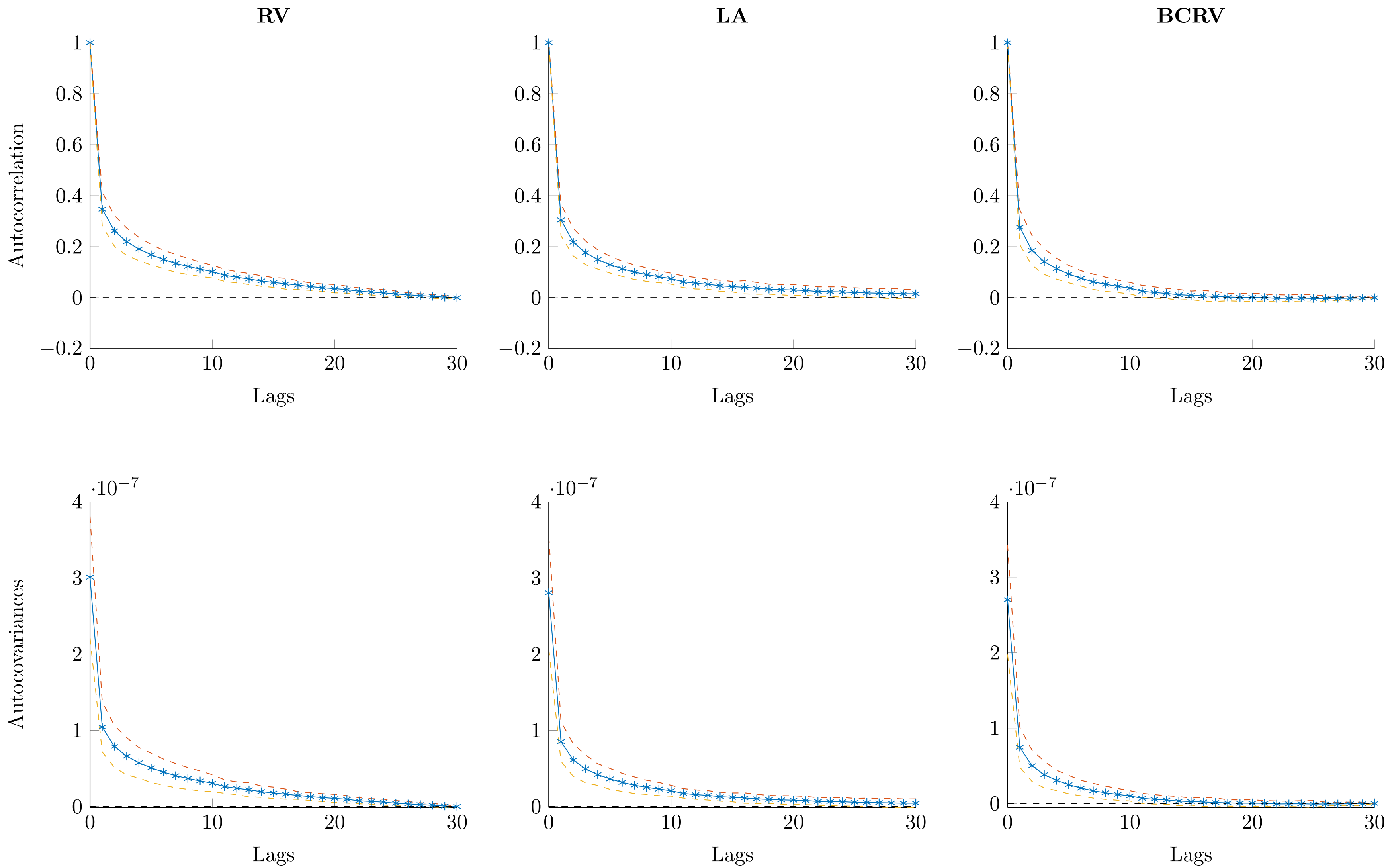

Figure 5: From the left to the right, we display the realized volatility (RV), local averaging (LA), and the bias corrected realized volatility (BCRV) estimators of the autocorrelations (top panel) and autocovariances (bottom panel) of noise against the number of lags j based on transaction data for Citigroup. Sample period: January, 2011. On average there are 10.5 observations per second in the sample. The three estimators are applied to and then averaged over each of the 20 trading days. The stars indicate the means of the 20 estimates. The dashed lines are 2 standard deviations away from the mean. The tuning parameter of the RV estimator is $j_n = 30$.

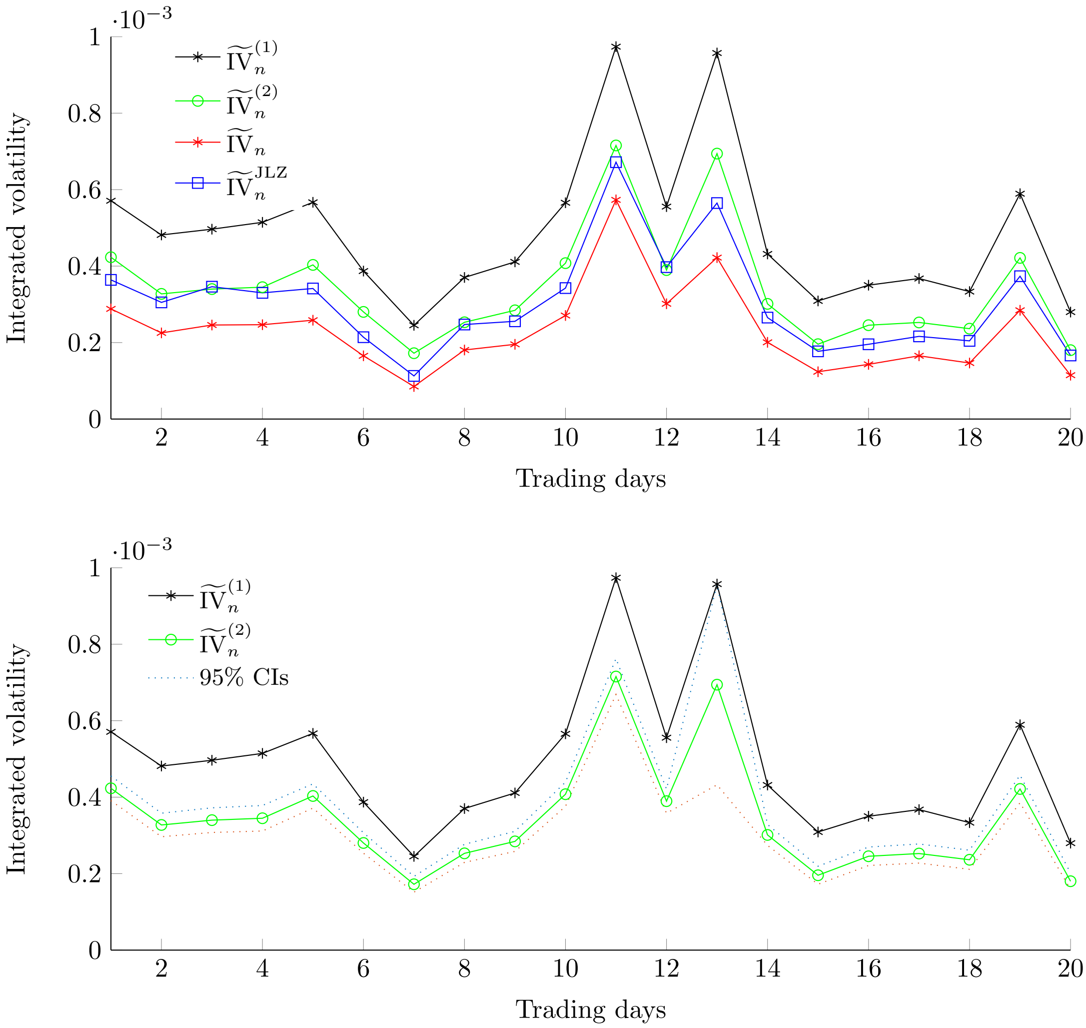

Figure 6: Estimation of the IV based on transaction data for Citigroup. Sample period: January, 2011, consisting of 20 trading days. On average there are 10.5 observations per second in the sample. The estimators $\widetilde{\mathrm{IV}}_n^{(1)}$, $\widetilde{\mathrm{IV}}_n^{(2)}$, and $\widetilde{\mathrm{IV}}_n$ are given by (35), (40), and (25). The $\widetilde{\mathrm{IV}}_n^{\mathrm{JLZ}}$ estimator is proposed in Jacod et al. (2019). In the bottom panel, the asymptotic confidence intervals (CIs) are based on the limit distribution in Theorem 5.1. The tuning parameter of the RV estimator is $j_n = 30$, and $\ell_n = 10$. θ is selected according to (28).

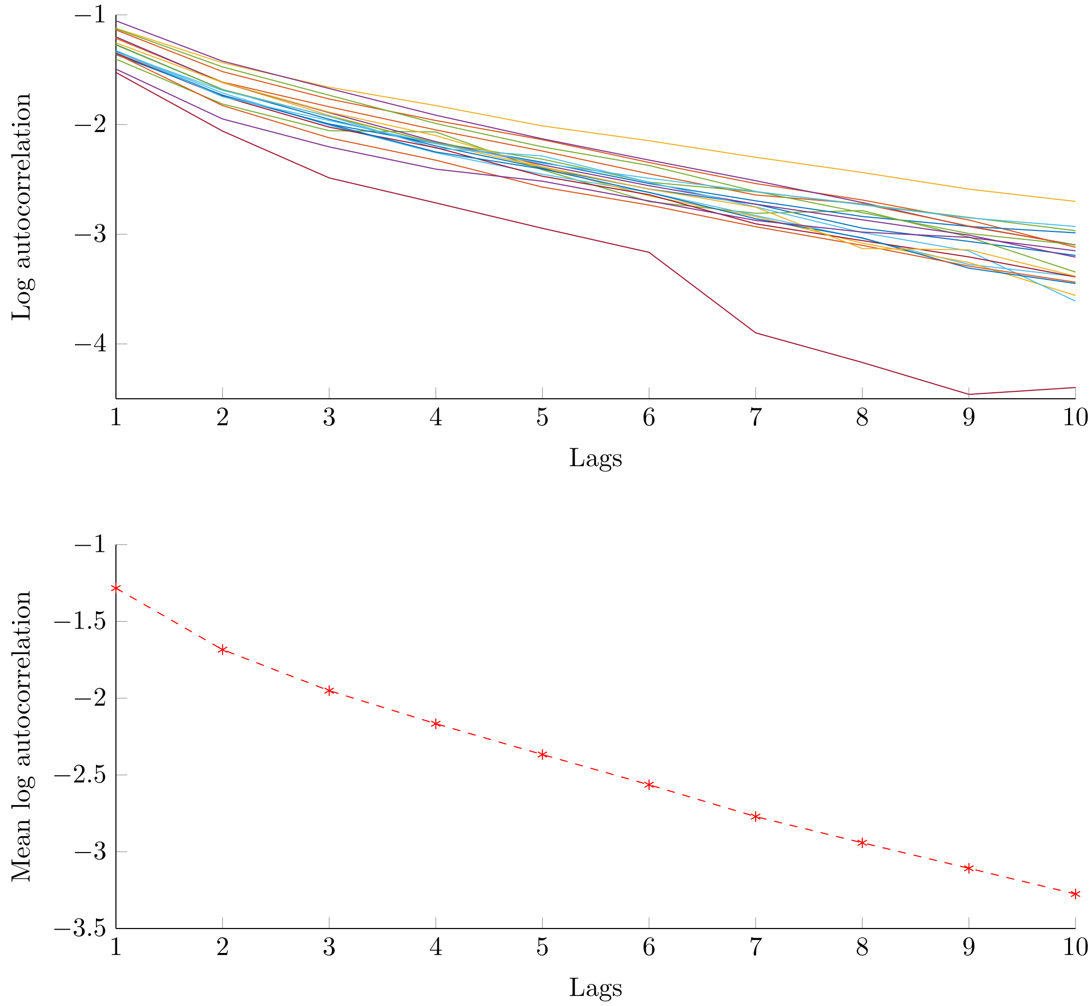

Figure 7: Top panel: Logarithmic autocorrelations of noise against the number of lags j estimated by BCRV for each trading day based on transaction data for Citigroup. Bottom panel: Means of the logarithmic autocorrelations of noise and a linear regression line. Sample period: January, 2011, consisting of 20 trading days. On average there are 10.5 observations per second in the sample. The tuning parameter of the RV estimator is $j_n = 30$.

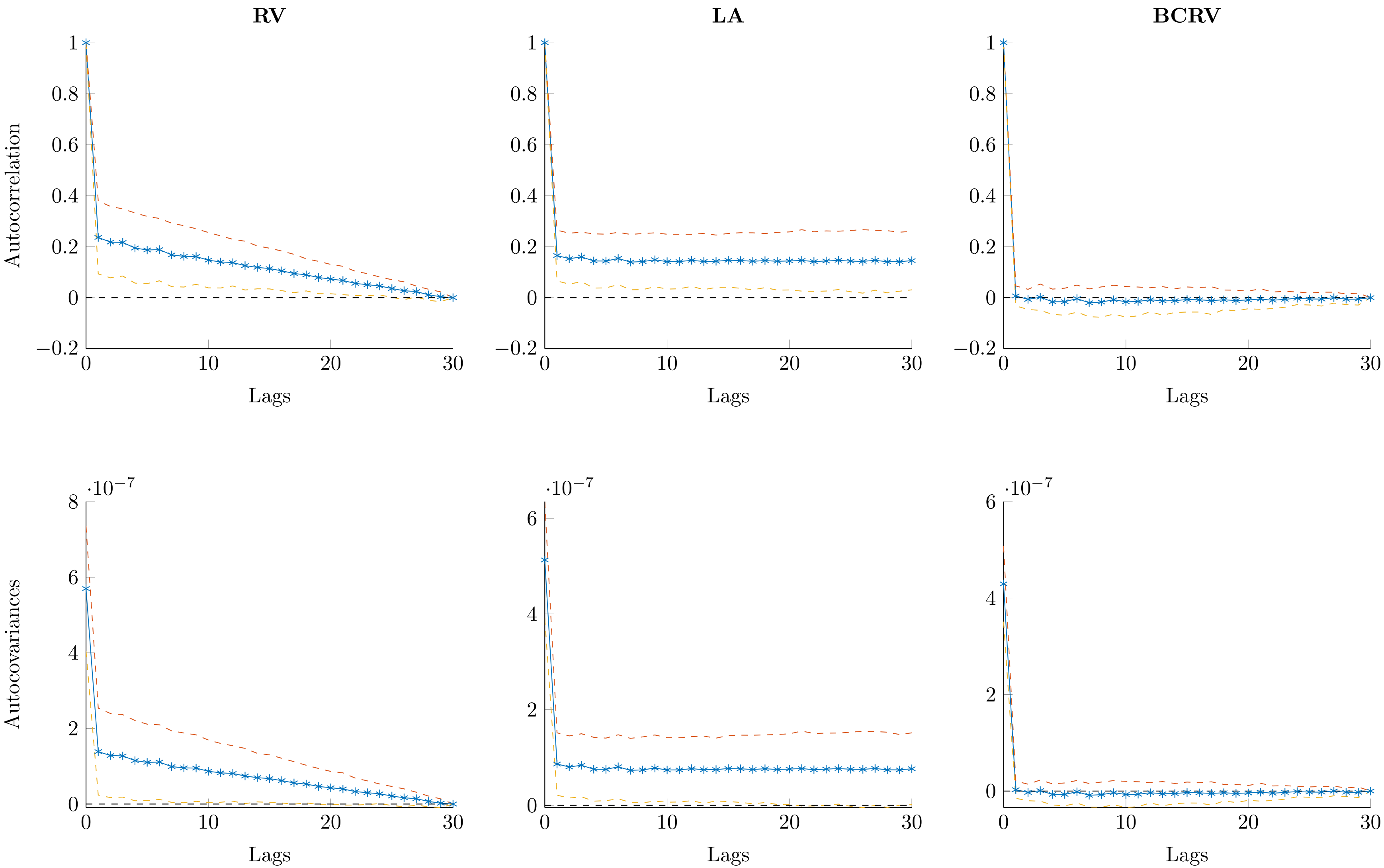

Figure 8: From the left to the right, we display the realized volatility (RV), local averaging (LA), and the bias corrected realized volatility (BCRV) estimators of the autocorrelations (top panel) and autocovariances (bottom panel) of noise against the number of lags j based on a subsample of the transaction data for Citigroup. Sample period: January, 2011. The subsample is recorded on a 1-sec time scale. The three estimators are applied to and next averaged over each of the 20 trading days. The stars indicate the means of the 20 estimates. The dashed lines are 2 standard deviations away from the mean. The tuning parameter of the RV estimator is $j_n = 30$.

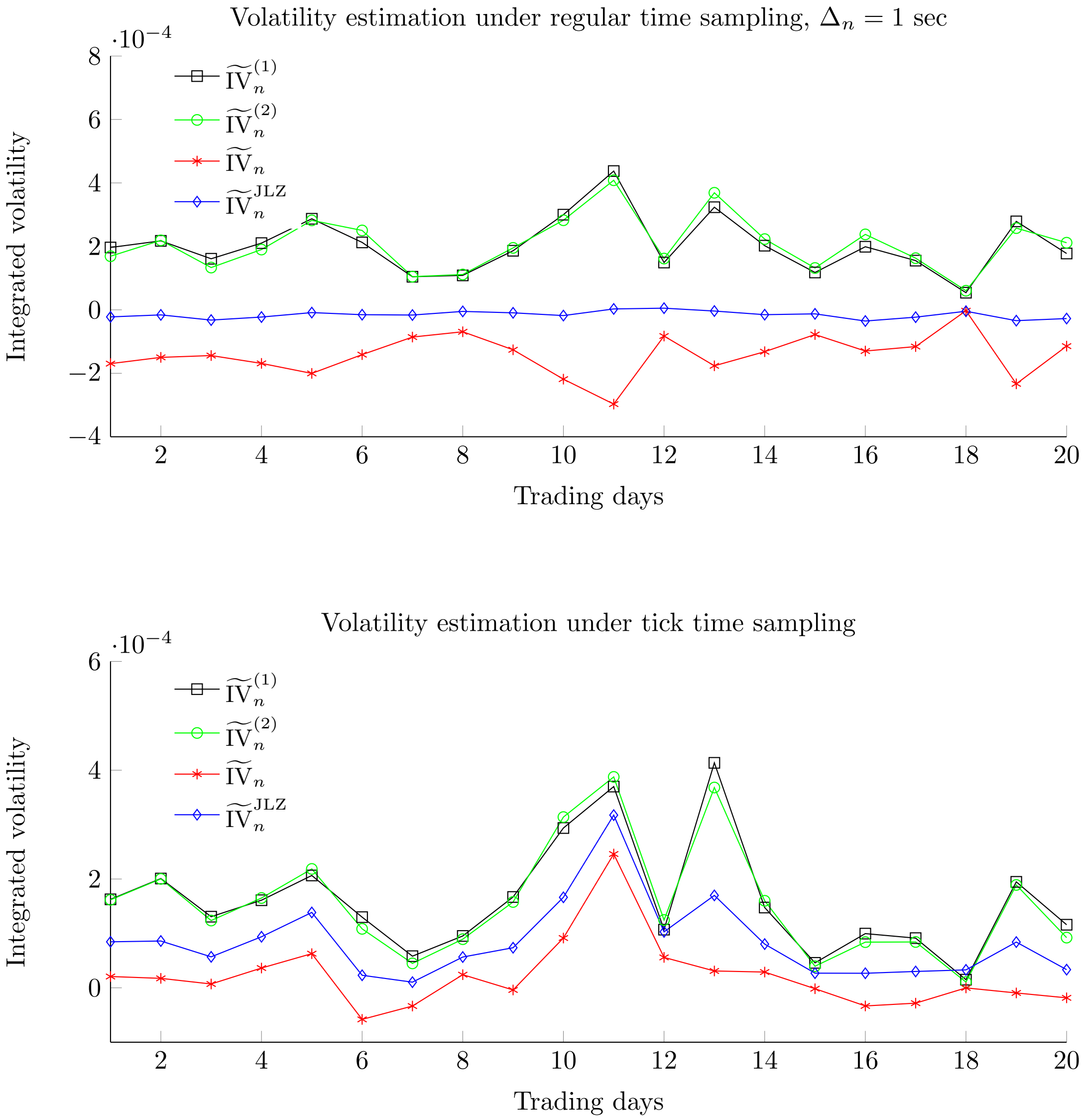

Figure 9: Estimation of the IV based on subsamples of the transaction data for Citigroup. Sample period: January, 2011, consisting of 20 trading days. In the top panel, the estimation is performed on a subsample that is recorded on a 1-sec time scale. In the bottom panel, the estimation is performed on a subsample that is recorded at tick time; on average there are 3.2 observations per second in the sample. The estimators $\widetilde{\mathrm{IV}}_n^{(1)}$, $\widetilde{\mathrm{IV}}_n^{(2)}$, and $\widetilde{\mathrm{IV}}_n$ are given by (35), (40), and (25). The $\widetilde{\mathrm{IV}}_n^{\mathrm{JLZ}}$ estimator is proposed in Jacod et al. (2019). The tuning parameter of the RV estimator is $j_n = 30$, and $\ell_n = 4$ for the 1-sec sample and $\ell_n = 6$ for the tick time sample. θ is selected according to (28).

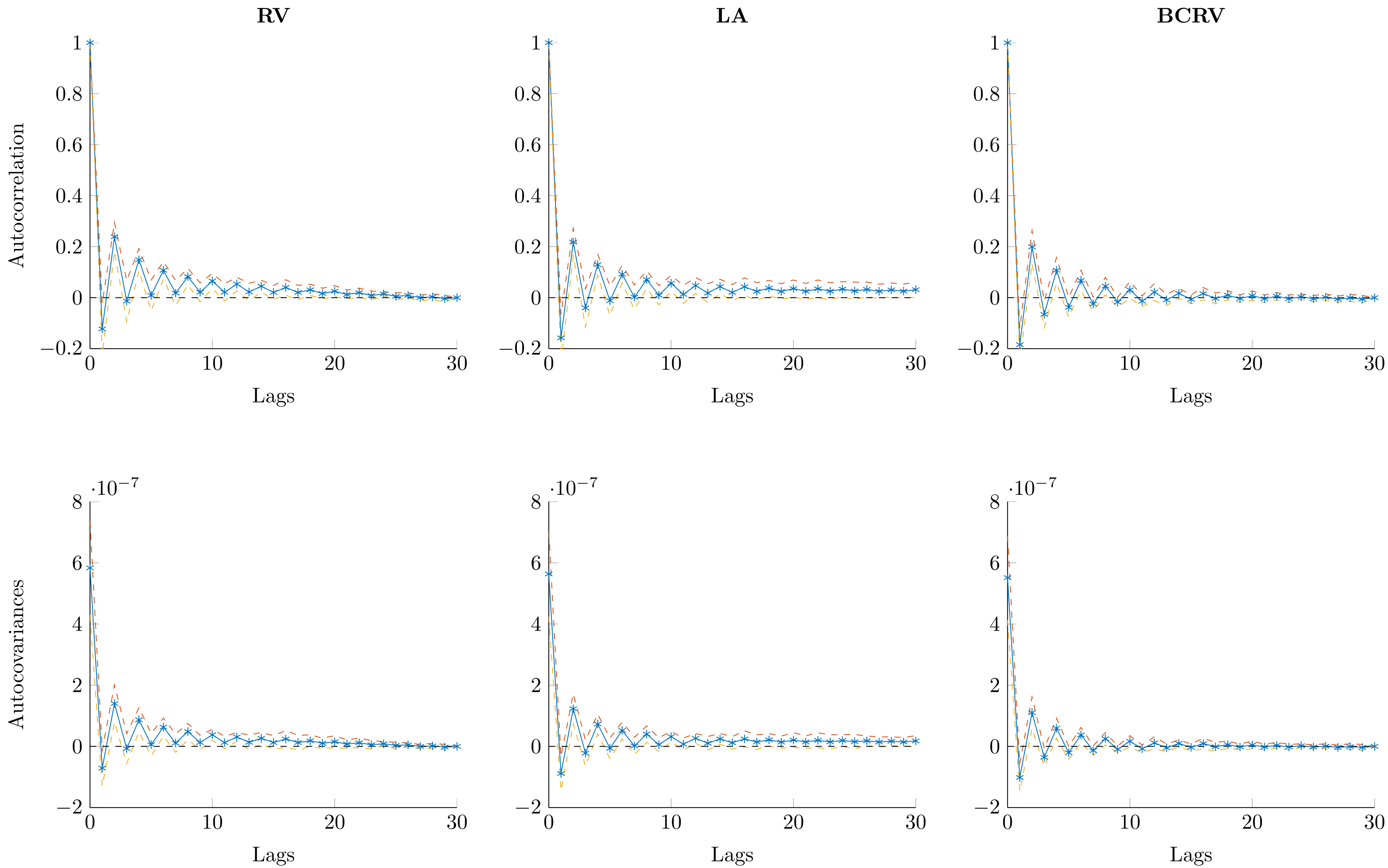

Figure 10: From the left to the right, we display the realized volatility (RV), local averaging (LA), and the bias corrected realized volatility (BCRV) estimators of the autocorrelations (top panel) and autocovariances (bottom panel) of noise against the number of lags j based on a subsample of the transaction data for Citigroup. Sample period: January, 2011. The subsample is recorded at tick time. On average there are 3.2 observations per second in the sample. The three estimators are applied to and then averaged over each of the 20 trading days. The stars indicate the means of the 20 estimates. The dashed lines are 2 standard deviations away from the mean. The tuning parameter of the RV estimator is $j_n = 30$.

Supplementary Material to "Dependent Microstructure Noise and Integrated Volatility Estimation from High-Frequency Data"

Z. Merrick Li*

Department of Economics

University of Cambridge

Roger J. A. Laeven†

Amsterdam School of Economics

University of Amsterdam, EURANDOM and CentER

Michel H. Vellekoop‡

Amsterdam School of Economics

University of Amsterdam

June 8, 2019

Abstract

Section A of this appendix contains detailed proofs of our results. In Sections B and C, we provide additional Monte Carlo simulation studies and empirical results.

*Corresponding author. Department of Economics, University of Cambridge, Austin Robinson Building, Sidgwick Avenue, Cambridge, CB3 9DD, United Kingdom. Email: Z.Merrick.Li@gmail.com. Phone: +44 (0) 1223 335288.

†University of Amsterdam, Amsterdam School of Economics, PO Box 15867, 1001 NJ Amsterdam, The Netherlands. Email: R.J.A.Laeven@uva.nl. Phone: +31 (0)20 5254219.

‡University of Amsterdam, Amsterdam School of Economics, PO Box 15867, 1001 NJ Amsterdam, The Netherlands. Email: M.H.Vellekoop@uva.nl. Phone: +31 (0)20 5254210.

A Proofs

A.1 Assumptions and Notation

In all proofs that follow, the constants C may vary from line to line, or even within one line. We add a subscript *par* if they depend on some parameter *par*. In the sequel, we will employ Lemma VIII.3.102 in Jacod and Shiryaev (2003) repeatedly, and we will refer to it as the *JS-Lemma*.

Adopting the standard localization procedure (see e.g., Jacod and Protter (2011) for further details), we may assume that:

Assumption A.1. *The efficient price X satisfies the Assumption 2.1 with b_t and σ_t bounded (uniformly in ω and t).*

This implies that for all stopping times $0 \leq S \leq T \leq 1$ we have

$$\begin{aligned}
&\mathbb{E}\left(|X_T - X_S|^p \,|\mathcal{F}_S\right) \leq C_p \mathbb{E}\left(T - S\,|\mathcal{F}_S\right), \quad \forall p \geq 2. \\
&\left|\mathbb{E}\left(X_T - X_S\,|\mathcal{F}_S\right)\right| \leq C\mathbb{E}\left(T - S\,|\mathcal{F}_S\right).
\end{aligned} \tag{A.1}$$

We first introduce some notation that is used to prove the results in Section 4.1:

$$\begin{aligned}
G_i^n(s) &:= \sum_{j=1}^{k_n-1} g_j^n \mathbf{1}_{\{((i+j-1)\Delta_n,(i+j)\Delta_n]\}}(s); \\
\mathcal{H}_i^n &:= \mathcal{F}_i^n \otimes \mathcal{G}_i; \\
\beta_m^n &:= n^{1/4}\left(\sigma_{\frac{m}{M_n}} \overline{W}_{mk_n}^n + \overline{U}_{mk_n}^n\right); \\
\xi_m^n &:= n^{1/4}\overline{Y}_{mk_n}^n - \beta_m^n; \\
\eta_m^n &:= \frac{n^{r/4}}{\theta}\mathbb{E}\left(\left|\overline{Y}_m^n\right|^r \left|\mathcal{H}_{mk_n}^n\right.\right); \\
\widetilde{\eta_m^n} &:= \frac{\mu_r}{\theta}\left(\theta\psi_0\sigma^2_{\frac{m}{M_n}} + \frac{\psi_1}{\theta}\Sigma_U\right)^{\frac{r}{2}}; \\
\mathrm{PAV}^n &:= \sum_{m=0}^{M_n-1} \eta_m^n; \\
\widetilde{\mathrm{PAV}}^n &:= \sum_{m=0}^{M_n-1} \widetilde{\eta_m^n}.
\end{aligned}$$

To prove the results presented in Section 4.2 we will also need the following:

$$\begin{aligned}
\widehat{G}_i^n(j,j') &= \int_0^\infty G_{i+j}^n(s)G_{i+j'}^n(s)\mathrm{d}s, \\
\overline{G}_i^n(j,j') &= \int_0^\infty G_{i+j}^n(s)G_{i+j'}^n(s)\mathrm{d}s \int_0^s G_{i+j}^n(u)G_{i+j'}^n(u)\mathrm{d}u, \\
X_i^n(t) &= B_i^n(t) + M_i^n(t);
\end{aligned}$$

where $B_i^n(t) = \int_0^t b_s G_i^n(s)\mathrm{d}s$, and $M_i^n(t) = \int_0^t \sigma_s G_i^n(s)\mathrm{d}W_s$. Furthermore, we define

$$\mathcal{K}_i^n = \mathcal{F}_i^n \otimes \mathcal{G}_{i-\left[\frac{k_n}{2}\right]}, \quad \mathcal{J}(p)_j^n = \mathcal{K}_{j(p+1)k_n}^n, \quad \mathcal{J}'(p)_j^n = \mathcal{K}_{j(p+1)k_n+pk_n}^n;$$
$$\widehat{c}_i^n = \sum_{j=1}^{k_n-1} \left(g_j^n\right)^2 \Delta_{i+j}^n C, \quad \alpha_n = \mathbb{E}\left(\left(\overline{U}_i^n\right)^2\right), \quad \widehat{U}_i^n = \left(\overline{U}_i^n\right)^2 - \alpha_n, \quad \widehat{X}_i^n = (\overline{X}_i^n)^2 - \widehat{c}_i^n;$$
$$\Psi_i^n = \left(\overline{Y}_i^n\right)^2 - \widehat{c}_i^n - \alpha_n = \widehat{X}_i^n + \widehat{U}_i^n + 2\overline{X}_i^n\overline{U}_i^n, \quad \zeta(p)_i^n = \sum_{j=i}^{i+pk_n-1} \Psi_j^n;$$
$$\eta(p)_j^n = \frac{\sqrt{\Delta_n}}{\theta\psi_0}\zeta(p)_{j(p+1)k_n}^n, \quad \overline{\eta}(p)_j^n = \mathbb{E}\left(\eta(p)_j^n \,|\mathcal{J}(p)_j^n\right);$$
$$\eta'(p)_j^n = \frac{\sqrt{\Delta_n}}{\theta\psi_0}\zeta(1)_{j(p+1)k_n+pk_n}^n, \quad \overline{\eta}'(p)_j^n = \mathbb{E}\left(\eta'(p)_j^n \,|\mathcal{J}'(p)_j^n\right),$$

and let $K_n^p = \lfloor \frac{1}{(p+1)k_n\Delta_n} \rfloor - 1, I_n^p = (K_n^p + 1)(p+1)k_n$. We can then decompose $\widetilde{\mathrm{IV}}_n - \mathrm{IV}$ into the following terms:

$$F(p)_n = \sum_{j=0}^{K_n^p} \overline{\eta}(p)_j^n, \quad M(p)_n = \sum_{j=0}^{K_n^p} \left(\eta(p)_j^n - \overline{\eta}(p)_j^n\right);$$
$$F'(p)_n = \sum_{j=0}^{K_n^p} \overline{\eta}'(p)_j^n, \quad M'(p)_n = \sum_{j=0}^{K_n^p} \left(\eta'(p)_j^n - \overline{\eta}'(p)_j^n\right);$$
$$\widehat{C}(p)_n = \frac{\sqrt{\Delta_n}}{\theta\psi_0} \sum_{i=I_n^p}^{n-k_n+1} \Psi_i^n;$$
$$\widehat{C}'(p)_n = \frac{(n-k_n+2)\,\alpha_n\sqrt{\Delta_n}}{\theta\psi_0} - \frac{\psi_1}{\theta^2\psi_0}\sum_{j=-\ell_n}^{\ell_n} \widehat{\gamma(j)}_n;$$
$$\widehat{C}''_n = \frac{\sqrt{\Delta_n}}{\theta\psi_0} \sum_{i=0}^{n-k_n+1} \widehat{c}_i^n - \mathrm{IV},$$

since we have

$$\widetilde{\mathrm{IV}}_n - \mathrm{IV} = M(p)_n + M'(p)_n + F(p)_n + F'(p)_n + \widehat{C}(p)_n + \widehat{C}'(p)_n + \widehat{C}''_n. \tag{A.2}$$

A.2 Auxiliary Lemmas

We will often need the following two useful results based on the JS-Lemma.

Let Z be an integrable random variable measurable with respect to $\mathcal{G}_{k'+k}$ (see Assumption 2.2 for the definition of this σ-algebra) and define

$$C_Z^k := \mathbb{E}\Big(\left(\mathbb{E}\left(Z\,|\mathcal{G}_{k'}\right) - \mathbb{E}(Z)\right)^2\Big), \quad \Lambda_Z := \frac{\mathbb{E}\left(Z\,|\mathcal{G}_{k'}\right) - \mathbb{E}(Z)}{\sqrt{C_Z^k}}.$$

Then we have by the JS-Lemma

$$\mathbb{E}\left(Z\,|\mathcal{G}_{k'}\right) = \mathbb{E}(Z) + \Lambda_Z\sqrt{C_Z^k}, \tag{A.3}$$

with $\mathbb{E}\big(\Lambda_Z^2\big) = 1$ and $C_Z^k \leq Ck^{-2v}$.

Another application of the JS-Lemma gives that if Z_i, Z_j are $\mathcal{G}_i$- and $\mathcal{G}_j$-measurable random variables respectively, with mean zero and bounded variance, then we have for all $k \leq i < j$ that

$$\mathbb{E}(|\mathbb{E}\left(Z_iZ_j\,|\mathcal{G}_k\right)|) \leq C\,(j-i)^{-v}\,. \tag{A.4}$$

To see this, we use the JS-Lemma to obtain (since the Z_j have bounded variance):

$$c_{ij} := \mathbb{E}\Big(\left(\mathbb{E}\left(Z_j\,|\mathcal{G}_i\right)\right)^2\Big) \leq C\,(j-i)^{-2v}\,. \tag{A.5}$$

Then,

$$\mathbb{E}(|\mathbb{E}\left(Z_iZ_j\,|\mathcal{G}_k\right)|) \leq \sqrt{C\,(j-i)^{-2v}}\,\mathbb{E}\left(\left|\mathbb{E}\left(Z_i\frac{\mathbb{E}\left(Z_j\,|\mathcal{G}_i\right)}{\sqrt{c_{ij}}}\,|\mathcal{G}_k\right)\right|\right).$$

Now applying the Cauchy-Schwarz inequality and using the fact that the variance of the Z_i is bounded, we obtain (A.4).

Next, in the setting of Section 4, we recall some useful estimates (see Jacod et al. (2009)) for pre-averaged sequences defined in (19):

$$\left|\mathbb{E}\left(\overline{X}_i^n\,|\mathcal{F}_i^n\right)\right| \leq C\sqrt{\Delta_n}, \quad \mathbb{E}\left(\left|\overline{X}_i^n\right|^q|\mathcal{F}_i^n\right) \leq C_q\Delta_n^{q/4}, \tag{A.6}$$

for any $q > 0$, and

$$\mathbb{E}\left(\left(\overline{W}_i^n\right)^2|\mathcal{F}_i^n\right) = k_n\Delta_n\psi_0 + O_p(\Delta_n^{3/4}). \tag{A.7}$$

The following lemma, which establishes a central limit theorem for general pre-averaged noise, plays a central role in the proofs of the results in Sections 4 and 5.

Lemma A.1. *Assume that the noise satisfies Assumption 2.2 and that (14) is satisfied. Then, the following central limit theorem holds for $\overline{U}_i^n$:*

$$n^{1/4}\overline{U}_i^n \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\psi_1\Sigma_U}{\theta}\right). \tag{A.8}$$

Proof. Let $a_j^n = -\overline{g}_j^n\sqrt{k_n/\phi_1^n(0)}$. First, a Riemann sum approximation implies

$$\phi_1^n(0) = \psi_1 + o(\Delta_n^{1/4}). \tag{A.9}$$

Next, for any $\ell \in \mathbb{Z}$, the Lipschitz property of g' implies $\left|\overline{g}_j^n - \overline{g}_{j-\ell}^n\right| \leq C|\ell|k_n^{-2}$, so

$$|\phi_1^n(\ell) - \psi_1| \leq C|\ell|/k_n + o(\Delta_n^{1/4}). \tag{A.10}$$

Since $\mathbb{E}\left(\left(\sum_{j=0}^{k_n-1} a_j^n U_{i+j}\right)^2\right) = \frac{1}{\phi_1^n(0)} \sum_{|\ell|\leq k_n} \phi_1^n(\ell)\gamma(\ell)$, we have

$$\left|\mathbb{E}\left(\left(\sum_{j=0}^{k_n-1} a_j^n U_{i+j}\right)^2\right) - \frac{1}{\phi_1^n(0)} \sum_{|\ell|\leq k_n} \psi_1\gamma(\ell)\right| \leq \frac{C}{\phi_1^n(0)} \sum_{|\ell|\leq k_n} \frac{|\gamma(\ell)\ell|}{k_n} \leq C\sqrt{\Delta_n}, \tag{A.11}$$

where we used that $|\gamma(\ell)\ell| \leq C|\ell|^{1-v}$ with $v > 2$, and $k_n = O(n^{1/2})$. Then $\sum_{|\ell|>k_n} \gamma(l) \leq C k_n^{1-v}$ gives

$$\left|\frac{1}{\phi_1^n(0)} \sum_{|\ell|\leq k_n} \psi_1\gamma(\ell) - \Sigma_U\right| \leq C\sqrt{\Delta_n}, \tag{A.12}$$

and we see that $\mathbb{E}\left(\left(\sum_{j=0}^{k_n-1} a_j^n U_{i+j}\right)^2\right) \to \Sigma_U$.

Since we assume the existence of moments of noise of all orders, and $v > 1$, we have for sufficiently large r that $v - \frac{2}{r-2} > 1$, which implies

$$\sum_{k\in\mathbb{N}^*} k^{\frac{2}{r-2}} \rho_k < \infty,$$

where the $\{\rho_k\}$ are the ρ-mixing coefficients. This is sufficient for the following CLT, according to Rio (1997)[1]:

$$\sum_{j=0}^{k_n-1} a_j^n U_{i+j} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \Sigma_U\right).$$

Since $n^{1/4}\overline{U}_i^n = \sqrt{\frac{\phi_1^n(0)}{\Delta_n^{1/2} k_n}} \sum_{j=0}^{k_n-1} a_j^n U_{i+j}$, we obtain by (A.11) and (A.12), using (18) and (A.9), that

$$\mathbb{E}\left(\left(n^{1/4}\overline{U}_i^n\right)^2\right) = \frac{\psi_1 \Sigma_U}{\theta} + o(\Delta_n^{1/4}), \tag{A.13}$$

and the stated result follows. □

The result of Lemma A.1 for the asymptotics of pre-averaged noise will allow us to prove the results in Subsection 4.1 using a similar strategy as in Podolskij and Vetter (2009a,b). However, their proofs need to be modified for our setting. The following lemmas will therefore turn out to be useful.

Lemma A.2. *Assume the conditions of Theorem 4.1 are satisfied. Then there is, for any $q > 0$, some*

[1] Rio (1997) discusses *strongly mixing* or α-*mixing*, which is implied by ρ-mixing.

constant $C_q > 0$ (depending on q), such that for all m:

$$\mathbb{E}(|\xi_m^n|^q) + \mathbb{E}\left(\left|n^{1/4}\overline{X}_i^n\right|^q\right) < C_q; \tag{A.14}$$

$$\mathbb{E}(|\beta_m^n|^q) + \mathbb{E}\left(\left|n^{1/4}\overline{Y}_i^n\right|^q\right) < C_q. \tag{A.15}$$

Proof of Lemma A.2. The boundedness of the moments of $n^{1/4}\overline{X}_i^n$ follows from (A.6), which also establishes the same bound for $n^{1/4}\overline{W}_i^n$ if we take the drift of X equal to zero and the volatility constant. This, together with the boundedness of σ, gives the boundedness of $\mathbb{E}(\xi_m^n)$ since we can write $\xi_m^n = n^{1/4}\left(\overline{X}_{mk_n}^n - \sigma_{m/M_n}\overline{W}_{mk_n}^n\right)$.

Now we show the boundedness of $\mathbb{E}\left(\left|n^{1/4}\overline{Y}_i^n\right|^q\right)$. Hölder's inequality implies

$$\mathbb{E}\left(\left|n^{1/4}\overline{Y}_i^n\right|^q\right) \le C_q\left(\mathbb{E}\left(\left|n^{1/4}\overline{X}_i^n\right|^q\right) + \mathbb{E}\left(\left|n^{1/4}\overline{U}_i^n\right|^q\right)\right).$$

Boundedness of $\mathbb{E}\left(\left|n^{1/4}\overline{X}_i^n\right|^q\right)$ has already been established, while $\mathbb{E}\left(\left|n^{1/4}\overline{U}_i^n\right|^q\right)$ is known to be bounded by Lemma A.1 and the well known fact that convergence in distribution implies convergence in moments under a uniformly bounded moments condition, see, e.g., Theorem 4.5.2 of Chung (2001). The result for $\mathbb{E}(|\beta_m^n|^q)$ follows by similar arguments. □

Lemma A.3. *Assume the conditions of Theorem 4.1 are satisfied. Then we have for all even integers $r > 2$ that, uniformly in m,*

$$\mathbb{E}\left((\beta_m^n)^2\,|\mathcal{H}_{mk_n}^n\right) = \left(\theta\psi_0\sigma^2_{\frac{m}{M_n}} + \frac{\psi_1\Sigma_U}{\theta}\right) + o_p(n^{-1/4}), \tag{A.16}$$

$$\mathbb{E}\left((\beta_m^n)^r\,|\mathcal{H}_{mk_n}^n\right) = \mu_r\left(\theta\psi_0\sigma^2_{\frac{m}{M_n}} + \frac{\psi_1\Sigma_U}{\theta}\right)^{r/2} + o_p(1), \tag{A.17}$$

with μ_r the moment of order r of a standard normal random variable.

Proof. Let $\{r_n\}$ be a sequence of integers satisfying

$$r_n \asymp n^{\vartheta}, \quad \frac{1}{4v} < \vartheta < \frac{1}{4}. \tag{A.18}$$

For any process Z, denote

$$\overline{Z}_{m,r_n}^n := -\sum_{j=0}^{r_n-1}\overline{g}_j^n Z_{mk_n+j}^n,$$

$$\overline{Z}_{r_n,m+1}^n := -\sum_{j=r_n}^{k_n-1}\overline{g}_j^n Z_{mk_n+j}^n.$$

Let

$$\overline{\beta}_{m,r_n}^n := n^{1/4}\overline{U}_{m,r_n}^n \qquad \overline{\beta}_{r_n,m+1}^n := n^{1/4}\left(\sigma_{\frac{m}{M_n}}\overline{W}_{mk_n}^n + \overline{U}_{r_n,m+1}^n\right). \tag{A.19}$$

This implies that $\beta_m^n = \overline{\beta}_{m,r_n}^n + \overline{\beta}_{r_n,m+1}^n$. We first prove (A.16) by establishing the following three results:

$$\mathbb{E}\left((\beta_m^n)^2 \,|\mathcal{H}_{mk_n}^n\right) - \mathbb{E}\left(\left(\overline{\beta}_{r_n,m+1}^n\right)^2 |\mathcal{H}_{mk_n}^n\right) = o_p(n^{-1/4}), \tag{A.20}$$

$$\mathbb{E}\left(\left(\overline{\beta}_{r_n,m+1}^n\right)^2 |\mathcal{H}_{mk_n}^n\right) - \mathbb{E}\left(\left(\overline{\beta}_{r_n,m+1}^n\right)^2 |\mathcal{F}_{mk_n}^n\right) = o_p(n^{-1/4}), \tag{A.21}$$

$$\mathbb{E}\left(\left(\overline{\beta}_{r_n,m+1}\right)^2 |\mathcal{F}_{mk_n}^n\right) - \left(\theta\psi_0\sigma^2_{\frac{m}{M_n}} + \frac{\psi_1\Sigma_U}{\theta}\right) = o_p(n^{-1/4}). \tag{A.22}$$

1. To prove (A.20), it is enough to show that

$$\mathbb{E}\left(\left(\overline{\beta}_{m,r_n}^n\right)^2 |\mathcal{H}_{mk_n}^n\right) = o_p(n^{-1/4}), \tag{A.23}$$

$$\mathbb{E}\left(\left(\overline{\beta}_{r_n,m+1}^n\right)\left(\overline{\beta}_{m,r_n}^n\right) |\mathcal{H}_{mk_n}^n\right) = o_p(n^{-1/4}). \tag{A.24}$$

To establish (A.23), we write

$$\left(\overline{\beta}_{m,r_n}^n\right)^2 = n^{1/2}\left(\overline{U}_{m,r_n}^n\right)^2 = n^{1/2}\sum_{j=0}^{r_n-1}\sum_{j'=0}^{r_n-1}\overline{g}_j^n\overline{g}_{j'}^n U_{mk_n+j}^n U_{mk_n+j'}^n.$$

Taking conditional expectations and using that $\overline{g}_j^n \leq C\sqrt{\Delta_n}$ for all j shows that the left-hand side in (A.23) is smaller than

$$C\sqrt{\Delta_n}\left(\sum_{j=0}^{r_n-1}\mathbb{E}\left((U_{mk_n+j})^2 \,|\mathcal{H}_{mk_n}^n\right) + 2\sum_{j=0}^{r_n-2}\sum_{j'=j+1}^{r_n-1}\mathbb{E}\left(U_{mk_n+j}U_{mk_n+j'} \,|\mathcal{H}_{mk_n}^n\right)\right)$$

so by (A.4)

$$\mathbb{E}\left(\left(\overline{\beta}_{m,r_n}^n\right)^2 |\mathcal{H}_{mk_n}^n\right) \leq \sqrt{\Delta_n}\left(Cr_n + 2\sum_{j=0}^{r_n-2}\sum_{j'=j+1}^{r_n-1}(j'-j)^{-v}\right) \;\leq\; C\sqrt{\Delta_n}r_n \tag{A.25}$$

and this proves (A.23) by (A.18). To prove (A.24) it is enough to show that

$$\sqrt{n}\mathbb{E}\left(\overline{U}_{m,r_n}^n\overline{U}_{r_n,m+1}^n \,|\mathcal{G}_{mk_n}\right) = o_p(n^{-1/2}), \tag{A.26}$$

$$n^{\frac{1}{4}}\mathbb{E}\Big(\mathbb{E}\left(\overline{\beta}_{m,r_n}^n\overline{W}_{mk_n}^n \,|\mathcal{H}_{mk_n}^n\right)\Big) = o_p(n^{-1/4}). \tag{A.27}$$

The first result follows since the left-hand side equals

$$\mathbb{E}\left(n^{1/2}\sum_{j=0}^{r_n-1}\sum_{j'=r_n}^{k_n-1}\overline{g}_j^n\overline{g}_{j'}^n U_{mk_n+j}^n U_{mk_n+j'}^n \mid \mathcal{G}_{mk_n}\right) \;\leq\; Cn^{1/2}\sum_{j=0}^{r_n-1}\sum_{j'=r_n}^{k_n-1}\sqrt{\Delta_n}\sqrt{\Delta_n}\,|j'-j|^{-v}$$

by (A.4), and since $v > 2$ we get (A.26).

For (A.27), we note that the independence of $\mathcal{G}, \mathcal{F}$, and the estimates (A.6) and (A.23) imply

$$n^{\frac{1}{4}}\mathbb{E}\Big(\Big|\mathbb{E}\left(\overline{\beta}^n_{m,r_n}\overline{W}^n_{mk_n}\left|\mathcal{H}^n_{mk_n}\right.\right)\Big|\Big) \le Cn^{\frac{1}{4}}\sqrt{\Delta_n}\mathbb{E}\Big(\Big|\mathbb{E}\left(\overline{\beta}^n_{m,r_n}\left|\mathcal{G}_{mk_n}\right.\right)\Big|\Big) \le C\Delta_n^{\frac{1}{4}}\sqrt{n^{-\frac{1}{4}}}. \tag{A.28}$$

This proves (A.24) and hence (A.20) has now been established.

2. To prove (A.21) we note that the left-hand side of (A.21) is

$$\mathbb{E}\left(\left(n^{\frac{1}{4}}\overline{U}^n_{r_n,m+1}\right)^2\left|\mathcal{H}^n_{mk_n}\right.\right) - \mathbb{E}\Big(\left(n^{\frac{1}{4}}\overline{U}^n_{r_n,m+1}\right)^2\Big),$$

which is of order $O_p(r_n^{-v})$ by (A.3), so (A.21) follows from (A.18).

3. Finally, we prove (A.22). We have, by (A.7) and since $m/M_n \ge (mk_n)\Delta_n$,

$$\mathbb{E}\left(\left(n^{1/4}\sigma_{\frac{m}{M_n}}\overline{W}^n_{mk_n}\right)^2\left|\mathcal{F}^n_{mk_n}\right.\right) = n^{1/2}\sigma^2_{\frac{m}{M_n}}k_n\Delta_n\psi_0 + o_p(n^{-1/4}) = \sigma^2_{\frac{m}{M_n}}\psi_0\theta + o_p(n^{-1/4}),$$

where the last equality follows from (18). Due to the independence of $\mathcal{G}$ and $\mathcal{F}$ we therefore only need to show that

$$\mathbb{E}\Big(\left(n^{1/4}\overline{U}^n_{r_n,m+1}\right)^2\Big) = \frac{\psi_1\Sigma_U}{\theta} + o_p(n^{-1/4}).$$

We know from (A.13) that

$$\mathbb{E}\Big(\left(n^{1/4}\overline{U}^n_{mk_n}\right)^2\Big) = \frac{\psi_1\Sigma_U}{\theta} + o(\Delta_n^{1/4}),$$

so the desired result follows if we can show that

$$\left|\mathbb{E}\Big(\left(n^{1/4}\overline{U}^n_{mk_n}\right)^2\Big) - \mathbb{E}\Big(\left(n^{1/4}\overline{U}^n_{r_n,m+1}\right)^2\Big)\right| = o_p(n^{-1/4}). \tag{A.29}$$

But this follows from

$$\mathbb{E}\Big(\left(n^{1/4}\overline{U}^n_{m,r_n}\right)^2\Big) \le C\Delta_n r_n; \quad \mathbb{E}\Big(\overline{U}^n_{m,r_n}\overline{U}^n_{r_n,m+1}\Big) \le C\Delta_n,$$

which can be obtained from (A.23) and (A.26).

This completes the proof of (A.16). To establish (A.17), we show that

$$\mathbb{E}\left(|\beta^n_m|^r\left|\mathcal{H}^n_{mk_n}\right.\right) - \mathbb{E}\left(|\overline{\beta}^n_{r_n,m+1}|^r\left|\mathcal{H}^n_{mk_n}\right.\right) = o_p(1), \tag{A.30}$$

$$\mathbb{E}\left(|\overline{\beta}^n_{r_n,m+1}|^r\left|\mathcal{H}^n_{mk_n}\right.\right) - \mathbb{E}\left(|\overline{\beta}^n_{r_n,m+1}|^r\left|\mathcal{F}^n_{mk_n}\right.\right) = o_p(n^{-1/4}), \tag{A.31}$$

$$\mathbb{E}\left(|\overline{\beta}_{r_n,m+1}|^r\left|\mathcal{F}^n_{mk_n}\right.\right) - \left(\theta\psi_0\sigma^2_{\frac{m}{M_n}} + \frac{\psi_1\Sigma_U}{\theta}\right)^{r/2} = o_p(1). \tag{A.32}$$

1. For (A.30), we use the Mean Value Theorem to write

$$\mathbb{E}\left((\beta^n_m)^r - \left(\overline{\beta}^n_{r_n,m+1}\right)^r |\mathcal{H}^n_{mk_n}\right) = \mathbb{E}\left(r\left(\overline{\beta}^n_{r_n,m+1}\right)^{r-1}\left(\overline{\beta}^n_{m,r_n}\right)|\mathcal{H}^n_{mk_n}\right) + o_p(1).$$

   Application of the Cauchy-Schwarz inequality yields that the right-hand side is $o_p(1)$ due to (A.25) and Lemma A.2.

2. We now turn to (A.31). For any $l \leq r$, apply (A.3) to write

$$\mathbb{E}\left(\left(n^{1/4}\overline{U}^n_{r_n,m}\right)^l |\mathcal{H}^n_{mk_n}\right) = \mathbb{E}\left(\left(n^{1/4}\overline{U}^n_{r_n,m}\right)^l\right) + C_{r_n,l}\Lambda_l,$$

   with $\mathbb{E}(\Lambda_l^2) = 1$ and $C_{r_n,l} \leq C r_n^{-v} \leq C n^{-1/4}$ because of (A.18). This means we can replace the conditional moments by the unconditional moments plus a correction term that vanishes asymptotically. Using the notation $C_r^k = \frac{r!}{k!(r-k)!}$ for the binomial coefficients, this gives:

$$\begin{aligned}
&\mathbb{E}\left(\left(\overline{\beta}^n_{r_n,m+1}\right)^r |\mathcal{H}^n_{mk_n}\right)\\
&= \mathbb{E}\left(\sum_{k=0}^r C_r^k \sigma^k_{\frac{m}{M_n}}\left(n^{1/4}\overline{W}^n_{mk_n}\right)^k \left(n^{1/4}\overline{U}^n_{r_n,m}\right)^{r-k} |\mathcal{H}^n_{mk_n}\right)\\
&= \sum_{k=0}^r C_r^k \sigma^k_{\frac{m}{M_n}} \mathbb{E}\left(\left(n^{1/4}\overline{W}^n_{mk_n}\right)^k |\mathcal{F}^n_{mk_n}\right)\mathbb{E}\left(\left(n^{1/4}\overline{U}^n_{r_n,m}\right)^{r-k} |\mathcal{G}_{mk_n}\right)\\
&= \mathbb{E}\left(\left(\overline{\beta}^n_{r_n,m+1}\right)^r |\mathcal{F}^n_{mk_n}\right) + \sum_{k=0}^r C_k^r \sigma^k_{\frac{m}{M_n}} \mathbb{E}\left(\left(n^{1/4}\overline{W}^n_{mk_n}\right)^k |\mathcal{F}^n_{mk_n}\right) C_{r_n,r-k}\Lambda_{r-k}.
\end{aligned}$$

   Clearly, the last term is $o_p(1)$ since (A.6) shows that the conditional expectation in the summation is bounded for all k, while $C_{r_n,l} \leq n^{-1/4}$. This proves (A.31).

3. The equality (A.32) is a consequence of the asymptotic distribution of β^n_m, which follows from Lemma A.1, the fact that the sequence of the moments of the noise is uniformly bounded, and the independence of W and U.

This concludes the proof of Lemma A.3. □

Lemma A.4. *Assume that the conditions of Theorem 4.2 hold and let*

$$L_n := n^{-1/4}\sum_{m=0}^{M_n-1}\left((\beta^n_m)^2 - \mathbb{E}\left((\beta^n_m)^2 |\mathcal{H}^n_{mk_n}\right)\right). \tag{A.33}$$

We have the following stable convergence in law:

$$L_n \xrightarrow{\mathcal{L}-s} \sqrt{\frac{2}{\theta}}\int_0^1 \left(\theta\psi_0\sigma_s^2 + \frac{\psi_1\Sigma_U}{\theta}\right)\mathrm{d}W'_s, \tag{A.34}$$

where W' is a standard Wiener process independent of $\mathcal{F}$.

Proof. Let $\vartheta_m^n := n^{-1/4}\left((\beta_m^n)^2 - \left(\theta\psi_0\sigma^2_{\frac{m}{M_n}} + \frac{\psi_1\Sigma_U}{\theta}\right)\right)$. Then, since $M_n \le C\sqrt{n}$,

$$L_n = \sum_{m=0}^{M_n-1} \vartheta_m^n + o_p(1),$$

by Lemma A.3. We also have

$$\sum_{m=0}^{M_n-1} \mathbb{E}\left(\vartheta_m^n \left|\mathcal{H}_{mk_n}^n\right.\right) \xrightarrow{\mathbb{P}} 0, \tag{A.35}$$

again by Lemma A.3, and

$$\begin{aligned}\sum_{m=0}^{M_n-1} \mathbb{E}\left((\vartheta_m^n)^2 \left|\mathcal{H}_{mk_n}^n\right.\right) =& \frac{1}{\theta M_n}\sum_{m=0}^{M_n-1} \mathbb{E}\left((\beta_m^n)^4 \left|\mathcal{H}_{mk_n}^n\right.\right) + \frac{1}{\theta M_n}\sum_{m=0}^{M_n-1}\left(\theta\psi_0\sigma^2_{\frac{m}{M_n}} + \frac{\psi_1\Sigma_U}{\theta}\right)^2 \\ &- \frac{2}{\theta M_n}\sum_{m=0}^{M_n-1} \mathbb{E}\left((\beta_m^n)^2 \left|\mathcal{H}_{mk_n}^n\right.\right)\left(\theta\psi_0\sigma^2_{\frac{m}{M_n}} + \frac{\psi_1\Sigma_U}{\theta}\right) + o_p(\Delta_n^{1/4}).\end{aligned}$$

The last remainder term $o_p(\Delta_n^{1/4})$ is due to the approximation $M_n = \sqrt{n}/\theta + o(n^{1/4})$. Now it follows from (A.60) and a Riemann approximation that

$$\sum_{m=0}^{M_n-1} \mathbb{E}\left((\vartheta_m^n)^2 \left|\mathcal{H}_{mk_n}^n\right.\right) \xrightarrow{\mathbb{P}} \frac{2}{\theta}\int_0^1 \left(\theta\psi_0\sigma^2_{\frac{m}{M_n}} + \frac{\psi_1\Sigma_U}{\theta}\right)^2 \mathrm{d}u. \tag{A.36}$$

Next, denote $\overline{\Delta_m^n Z} = Z^n_{(m+1)k_n} - Z^n_{mk_n}$, for any process Z. We will show that

$$\sum_{m=0}^{M_n-1} \mathbb{E}\left(\vartheta_m^n \overline{\Delta_m^n N} \left|\mathcal{H}_{mk_n}^n\right.\right) \xrightarrow{\mathbb{P}} 0, \tag{A.37}$$

for any bounded martingale N defined on the probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\ge 0}, \mathbb{P})$.

According to Jacod et al. (2009) and the proof of Theorem IX 7.28 of Jacod and Shiryaev (2003) it suffices to consider martingales in $\mathcal{N}^0$ or $\mathcal{N}^1$, where $\mathcal{N}^0$ is the set of all bounded martingales on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\ge 0}, \mathbb{P})$ which are orthogonal to W, and where $\mathcal{N}^1$ is the set of all martingales having a limit $N_\infty = f(Y_{t_1}, \dots, Y_{t_q})$, where f is any bounded Borel function on $\mathbb{R}^q$, $t_1 < \dots < t_q$ and $q \ge 1$.

First, let $N \in \mathcal{N}^0$ and let $\widetilde{\mathcal{F}}'_t = \bigcap_{s>t}\mathcal{F}_s \otimes \mathcal{G}$. Then, for any $t > \frac{m}{M_n}$, $\overline{\vartheta}_m^n(t) := \mathbb{E}\left(\vartheta_m^n \middle| \widetilde{\mathcal{F}}'_t\right)$, conditional on $\sigma_{\frac{m}{M_n}}$, is a martingale with respect to the filtration generated by $\{W_t - W_{\frac{m}{M_n}} | t > \frac{m}{M_n}\}$. By the martingale representation theorem, we have $\overline{\vartheta}_m^n(t) = \overline{\vartheta}_m^n(\frac{m}{M_n}) + \int_{\frac{m}{M_n}}^t \tau_u \mathrm{d}W_u$ for some predictable process τ. The orthogonality of W and N and the martingale property of N imply that

$$\mathbb{E}\left(\vartheta_m^n \overline{\Delta_m^n N} \middle| \widetilde{\mathcal{F}}'_{\frac{m}{M_n}}\right) = \mathbb{E}\left(\left(\vartheta_m^n - \overline{\vartheta}_m^n\left(\frac{m}{M_n}\right)\right)\overline{\Delta_m^n N} + \overline{\vartheta}_m^n\left(\frac{m}{M_n}\right)\overline{\Delta_m^n N} \middle| \widetilde{\mathcal{F}}'_{\frac{m}{M_n}}\right) = 0,$$

which gives

$$\mathbb{E}\left(\vartheta_m^n \overline{\Delta_m^n N}\,\middle|\mathcal{H}_{mk_n}^n\right) = 0, \tag{A.38}$$

since $\mathcal{H}_{mk_n}^n \subset \widetilde{\mathcal{F}}'_{\frac{m}{M_n}}$.

Next, assume that $N \in \mathcal{N}^1$. It can be shown (see Jacod et al. (2009)) that there exists some $\hat{f}_t$ such that $t \in [t_l, t_{l+1})$, $N_t = \hat{f}_t(Y_{t_0}, Y_{t_1}, \ldots, Y_{t_l})$ with $t_0 = 0, t_{q+1} = \infty$, and such that it is measurable in $(Y_{t_1}, \ldots, Y_{t_l})$. Hence, $\overline{\Delta_m^n N} = 0$ if it does not cover any of the points $t_1, \ldots, t_{q+1}$. But such intervals (to compute $\overline{\Delta_m^n N}$) that contain any of $t_1, \ldots, t_{q+1}$ are at most finite in number. Furthermore, by the boundedness of N and the conditional Cauchy-Schwarz inequality, we have the following:

$$\mathbb{E}\left(\left|\vartheta_m^n \overline{\Delta_m^n N}\right| \middle|\mathcal{H}_{mk_n}^n\right) \leq \sqrt{\mathbb{E}\left((\vartheta_m^n)^2 \middle|\mathcal{H}_{mk_n}^n\right)}\sqrt{\mathbb{E}\left(\left(\overline{\Delta_m^n N}\right)^2 \middle|\mathcal{H}_{mk_n}^n\right)} = O_p(n^{-1/4}).$$

Now (A.37) follows since there are at most finitely many such intervals.

Due to the fact that ϑ_m^n is an even functional of $n^{1/4}\overline{W}_{mk_n}^n$ and $n^{1/4}\overline{U}_{mk_n}^n$ we know that both have a symmetric asymptotic distribution, and

$$\mathbb{E}\left(\vartheta_m^n \overline{\Delta_m^n W}\,\middle|\mathcal{H}_{mk_n}^n\right) \xrightarrow{\mathbb{P}} 0. \tag{A.39}$$

From (A.17), we deduce that $(\vartheta_m^n)^2 \mathbf{1}_{\{|\vartheta_m^n| > \varepsilon\}} = o_p(n^{-1/2})$ for any $\varepsilon > 0$, so we have

$$\sum_{m=0}^{M_n-1} \mathbb{E}\left((\vartheta_m^n)^2 \mathbf{1}_{\{|\vartheta_m^n| > \varepsilon\}} \middle|\mathcal{H}_{mk_n}^n\right) \xrightarrow{\mathbb{P}} 0. \tag{A.40}$$

Now the proof is complete in view of (A.35)-(A.40), and Theorem IX.7.28 of Jacod and Shiryaev (2003). □

Lemma A.5. *Assume that the conditions of Theorem 4.2 hold. We then have that*

$$\sum_{m=0}^{M_n-1} \left(\overline{Y}_{mk_n}^n\right)^2 - \frac{1}{\sqrt{n}} \sum_{m=0}^{M_n-1} (\beta_m^n)^2 = o_p(n^{-1/4}). \tag{A.41}$$

Proof. Denote

$$\widetilde{Y}_m^n = n^{-1/4}\beta_m^n = \sigma_{\frac{m}{M_n}} \overline{W}_{mk_n}^n + \overline{U}_{mk_n}^n. \tag{A.42}$$

Then,

$$\mathbb{E}\left(\left|\sum_{m=0}^{M_n-1} \left(\overline{Y}_{mk_n}^n\right)^2 - \frac{1}{\sqrt{n}} \sum_{m=0}^{M_n-1} (\beta_m^n)^2\right|\right) \leq \sum_{m=0}^{M_n-1} \sqrt{\mathbb{E}\left(\left(\overline{Y}_{mk_n}^n - \widetilde{Y}_m^n\right)^2\right)}\sqrt{\mathbb{E}\left(\left(\overline{Y}_{mk_n}^n + \widetilde{Y}_m^n\right)^2\right)}.$$

Since $\sqrt{\mathbb{E}\left(\left(\overline{Y}^n_{mk_n} + \widetilde{Y}^n_m\right)^2\right)} = O(n^{-1/4})$ by (A.6), the result is proven if

$$\sum_{m=0}^{M_n-1} \sqrt{\mathbb{E}\left(\left(\overline{Y}^n_{mk_n} - \widetilde{Y}^n_m\right)^2\right)} \to 0. \tag{A.43}$$

But this follows directly from Lemma 7.8 in Barndorff-Nielsen et al. (2006). □

A.3 Proofs of the Results in Section 3 and Subsection 4.1

A.3.1 Proof of Proposition 3.1

Proof. For any process Z, we write $\Delta^n_{i,j} Z := Z^n_{i+j} - Z^n_i$, for $j = 1, 2, \ldots, n-i$. The process Y then satisfies

$$\sum_{i=0}^{n-j} (\Delta^n_{i,j} Y)^2 = \sum_{i=0}^{n-j} (\Delta^n_{i,j} X)^2 + 2 \sum_{i=0}^{n-j} \Delta^n_{i,j} X \; \Delta^n_{i,j} U + \sum_{i=0}^{n-j} (\Delta^n_{i,j} U)^2. \tag{A.44}$$

We now analyze the asymptotic properties of the three components on the right-hand side of (A.44):

(i) First note that $\sum_{i=0}^{n-j} (\Delta^n_{i,j} X)^2 / j \xrightarrow{\mathbb{P}} [X, X]$, where $[X, X]$ is the quadratic variation of X.

(ii) By the independence of X and U, we have

$$\sum_{i=0}^{n-j} \mathbb{E}\left(\left(\Delta^n_{i,j} X \; \Delta^n_{i,j} U\right)^2\right) = \sum_{i=0}^{n-j} \mathbb{E}\left(\left(\Delta^n_{i,j} X\right)^2\right) \mathbb{E}\left(\left(\Delta^n_{i,j} U\right)^2\right) \le Cj. \tag{A.45}$$

The last inequality follows from the fact that U has bounded moments and from an application of (A.1). Next,

$$\begin{aligned}
&\sum_{i,i':i<i'} \mathbb{E}\big(\Delta^n_{i,j} X \; \Delta^n_{i,j} U \; \Delta^n_{i',j} X \; \Delta^n_{i',j} U\big) \\
=& \sum_{i,i':i<i'} \mathbb{E}\big(\Delta^n_{i,j} X \; \Delta^n_{i',j} X\big) \mathbb{E}\big(\Delta^n_{i,j} U \; \Delta^n_{i',j} U\big) \\
\le& Cj\Delta_n \left(\sum_{i,i':i+j<i'} \mathbb{E}\big(\Delta^n_{i,j} U \; \Delta^n_{i',j} U\big) + \sum_{i,i':i+j \ge i' > i} \mathbb{E}\big(\Delta^n_{i,j} U \; \Delta^n_{i',j} U\big) \right) \\
\le& Cj^2.
\end{aligned} \tag{A.46}$$

The first inequality follows from the Cauchy-Schwarz inequality and (A.1). To establish the second inequality, we apply the Cauchy-Schwarz inequality, (A.5), and the fact that $v > 1$, to obtain

$$\begin{aligned}
\sum_{i,i':i+j<i'} \mathbb{E}\big(\Delta^n_{i,j} U \; \Delta^n_{i',j} U\big) &= \sum_{i,i':i+j<i'} \mathbb{E}\big(\Delta^n_{i,j} U \; \mathbb{E}\left(\Delta^n_{i',j} U \,|\mathcal{F}_{(i+j)\Delta_n}\right)\big) \\
&\le C \sum_i \sum_{i':i+j<i'} \sqrt{\mathbb{E}\left(\left(\mathbb{E}\left(\Delta^n_{i',j} U \,|\mathcal{F}_{(i+j)\Delta_n}\right)\right)^2\right)} \\
&\le C \sum_i \sum_{i':i+j<i'} (i' - (i+j))^{-v} \le C\Delta_n^{-1}.
\end{aligned} \tag{A.47}$$

Equations (A.45) and (A.46) imply that $\mathbb{E}\left(\left(\sum_{i=0}^{n-j}\Delta_{i,j}^n X\ \Delta_{i,j}^n U\right)^2\right)\le Cj^2$, so

$$\sum_{i=0}^{n-j}\Delta_{i,j}^n X\ \Delta_{i,j}^n U = O_p(j). \tag{A.48}$$

(iii) Turning to the last sum of (A.44), let $\nu_j := \mathbb{E}\big((U_{i+j}^n - U_i^n)^2\big) = 2(\gamma(0)-\gamma(j))$. For $i>j$, we obtain the following, using similar arguments as the ones used to prove (A.47):

$$\left|\mathbf{Cov}\big((U_j^n-U_0^n)^2,(U_{i+j}^n-U_i^n)^2\big)\right|\le C(i-j)^{-v},$$

which implies

$$\mathbb{E}\left(\left(\sum_{i=0}^{n-j}\left((\Delta_{i,j}^n U)^2-\nu_j\right)\right)^2\right)\le C\Delta_n^{-1}j. \tag{A.49}$$

For any fixed j and any j_n satisfying $\Delta_n j_n\to 0, j_n\to\infty$, we have by (A.48), (A.49) and (4) that

$$\begin{aligned}\widehat{\langle Y,Y\rangle}(j)_n-(\gamma(0)-\gamma(j)) &= O_p\left(\sqrt{\Delta_n j}\right);\\ \widehat{\langle Y,Y\rangle}(j_n)_n-\gamma(0) &= O_p\left(\max\left\{\sqrt{\Delta_n j_n}, j_n^{-v}\right\}\right).\end{aligned} \tag{A.50}$$

Now the stated result follows. □

A.3.2 Proof of Proposition 3.2

Proof. By Itô's isometry, we have

$$\begin{aligned}\mathbb{E}_\sigma\left(\sum_{i=0}^{n-j}(\Delta_{i,j}^n X)^2\right) &= \sum_{i=0}^{n-j}\sum_{h=i}^{i+j-1}\int_{h\Delta n}^{(h+1)\Delta_n}\sigma_s^2\mathrm{d}s = \sum_{h=0}^{n-1}\sum_{i=0\vee(h-j+1)}^{(n-j)\wedge h}\int_{h\Delta n}^{(h+1)\Delta_n}\sigma_s^2\mathrm{d}s\\ &= \sum_{h=j-1}^{n-j}\sum_{i=h-j+1}^{h}\int_{h\Delta n}^{(h+1)\Delta_n}\sigma_s^2\mathrm{d}s+o(j^2\Delta_n) = j\int_{(j-1)\Delta_n}^{(n+1-j)\Delta_n}\sigma_s^2\mathrm{d}s+o(j^2\Delta_n),\end{aligned}$$

where we reversed the order of summation in the second equality, while the stochastic orders follow from the regularity conditions on the volatility path around 0 and 1. Hence, we have

$$2(n-j+1)\mathbb{E}_\sigma\left(\widehat{\langle X,X\rangle}(j)_n\right) = j\int_0^1\sigma_s^2\mathrm{d}s+O_p(j^2\Delta_n).$$

Furthermore, it is immediate that $\mathbb{E}_\sigma\left(\sum_{i=0}^{n-j}(\Delta_{i,j}^n U)^2\right) = 2(n-j+1)(\gamma(0)-\gamma(j))$. Thus, we have, by the independence of X and U,

$$\mathbb{E}_\sigma\left(\widehat{\langle Y,Y\rangle}(j)_n\right) = \frac{j\int_0^1\sigma_s^2\mathrm{d}s}{2(n-j+1)}+\gamma(0)-\gamma(j)+O_p(j^2\Delta_n^2).$$

□

A.3.3 Proof of Proposition 3.3

Proof. We note that $|\Sigma_U - \widehat{\Sigma}_{U_n}|$ is smaller than

$$2\sum_{j=0}^{\ell_n} |\gamma(j) - \widehat{\gamma(j)}_n| + 2\sum_{j=\ell_n+1}^{\infty} |\gamma(j)|.$$

The last sum is of order $(\ell_n)^{1-v}$ with $v > 3$ and $\ell_n \geq Cn^{1/8}$ by (14), so it is $o(n^{-1/4})$. For the first sum we use (A.50) to conclude that for $j \leq \ell_n$:

$$|\gamma(j) - \widehat{\gamma(j)}_n| = |\gamma(j) - \widehat{\langle Y, Y\rangle}(j_n)_n + \widehat{\langle Y, Y\rangle}(j)_n| = O_p\left(\max\left\{\sqrt{\Delta_n j_n}, j_n^{-v}, \sqrt{\Delta_n \ell_n}\right\}\right).$$

Our restrictions in (14) then guarantee that $|\gamma(j) - \widehat{\gamma(j)}_n| = O_p((\Delta_n)^{5/12})$ while $\ell_n = o((\Delta_n)^{-1/6})$, so

$$\sum_{j=0}^{\ell_n} |\gamma(j) - \widehat{\gamma(j)}_n| = o_p(n^{-1/4}); \qquad |\Sigma_U - \widehat{\Sigma}_{U_n}| = o_p(n^{-1/4}). \tag{A.51}$$

This establishes the result. $\square$

A.3.4 Proof of Theorem 4.1

Proof. We present the proof in three steps, which correspond to the following three equations:

$$\mathrm{PAV}(Y, r)_n - \frac{1}{M_n}\mathrm{PAV}^n \xrightarrow{\mathbb{P}} 0, \tag{A.52}$$

$$\frac{1}{M_n}\mathrm{PAV}^n - \frac{1}{M_n}\widetilde{\mathrm{PAV}}^n \xrightarrow{\mathbb{P}} 0, \tag{A.53}$$

$$\frac{1}{M_n}\widetilde{\mathrm{PAV}}^n - \mathrm{PAV}(Y, r) \xrightarrow{\mathbb{P}} 0. \tag{A.54}$$

We invoke Lemmas A.2 and A.3, which in turn rely on Lemma A.1.

(i) To prove (A.52), recall our choice[2] of $M_n = \left\lfloor \frac{n}{k_n} \right\rfloor$. The difference on the left-hand side of (A.52) is a sum of martingale differences:

$$\mathrm{PAV}(Y, r)_n - \frac{1}{M_n}\mathrm{PAV}^n$$
$$= \sum_{m=0}^{M_n-1} \frac{1}{\sqrt{n}}\left(\left|n^{\frac{1}{4}}\overline{Y}^n_{mk_n}\right|^r - \mathbb{E}\left(\left|n^{\frac{1}{4}}\overline{Y}^n_{mk_n}\right|^r \Big| \mathcal{H}^n_{mk_n}\right)\right).$$

In light of Lemma 2.2.11 in Jacod and Protter (2011), it suffices to show that

$$\frac{1}{n}\sum_{m=0}^{M_n-1} \mathbb{E}\left(\left|n^{\frac{1}{4}}\overline{Y}^n_{mk_n}\right|^{2r} \Big| \mathcal{H}^n_{mk_n}\right) \xrightarrow{\mathbb{P}} 0. \tag{A.55}$$

But this follows from the boundedness established in Lemma A.2 and the choice of M_n.

[2] We interchangeably use $k_n\Delta_n$ and $1/M_n$ in the sequel; the difference of the two is always negligible.

(ii) To establish (A.53), we proceed in several steps:

(a) We first note that the error when approximating $n^{1/4}\overline{Y}_i^n$ by β_m^n, denoted by ξ_m^n, is small in the sense that

$$\frac{1}{M_n}\sum_{m=0}^{M_n-1}\mathbb{E}\Big(|\xi_m^n|^2\Big)\to 0. \tag{A.56}$$

To see this, we write

$$\xi_m^n = n^{1/4}\left(\int_{mk_n\Delta_n}^{(m+1)k_n\Delta_n} b_s G_{mk_n}^n(s)\mathrm{d}s + \int_{mk_n\Delta_n}^{(m+1)k_n\Delta_n}\left(\sigma_s - \sigma_{\frac{m}{M_n}}\right)G_{mk_n}^n(s)\mathrm{d}W_s\right).$$

Since b and G are bounded, we have

$$\mathbb{E}\left(n^{1/2}\left(\int_{mk_n\Delta_n}^{(m+1)k_n\Delta_n} b_s G_{mk_n}^n(s)\mathrm{d}s\right)^2\right) \le Cn^{1/2}(k_n\Delta_n)^2 \le C\sqrt{\Delta_n}.$$

By Itô isometry,

$$\begin{aligned}&\mathbb{E}\left(n^{1/2}\left(\int_{mk_n\Delta_n}^{(m+1)k_n\Delta_n}\left(\sigma_s - \sigma_{\frac{m}{M_n}}\right)G_{mk_n}^n(s)\mathrm{d}W_s\right)^2\right)\\ &\qquad \le\; C\Delta_n^{-1/2}\int_{mk_n\Delta_n}^{(m+1)k_n\Delta_n}\mathbb{E}\left(\left(\sigma_s - \sigma_{\frac{m}{M_n}}\right)^2\right)\mathrm{d}s,\end{aligned}$$

and hence

$$\frac{1}{M_n}\sum_{m=0}^{M_n-1}\mathbb{E}\Big(|\xi_m^n|^2\Big) \le C\left(\Delta_n^{1/2} + \int_0^1 \mathbb{E}\left(\left(\sigma_s - \sigma_{\frac{\lfloor M_n s\rfloor}{M_n}}\right)^2\right)\mathrm{d}s\right)\to 0,$$

by Lebesgue's dominated convergence theorem, since $\sigma_{\frac{\lfloor M_n s\rfloor}{M_n}} \to \sigma_s$ and σ is bounded.

(b) Next, define the approximation error

$$\zeta_m^n := \frac{\left|n^{1/4}\overline{Y}_{mk_n}^n\right|^r - |\beta_m^n|^r}{\theta}.$$

We note that this error is also small:

$$\frac{1}{M_n}\sum_{m=0}^{M_n-1}\mathbb{E}(|\zeta_m^n|)\to 0, \tag{A.57}$$

which follows from

$$\frac{1}{M_n}\sum_{m=0}^{M_n-1}\mathbb{E}\Big(|\zeta_m^n|^2\Big)\to 0. \tag{A.58}$$

This can be proved using similar arguments as in the proof of (A.56). Equation (A.57) then follows, and it implies

$$\frac{1}{M_n}\sum_{m=0}^{M_n-1}\mathbb{E}\left(\zeta_m^n\left|\mathcal{H}_{mk_n}^n\right.\right)\xrightarrow{\mathbb{P}}0, \tag{A.59}$$

by the Markov inequality.

(c) By Lemma A.3 we have

$$\mathbb{E}\left(|\beta_m^n|^r\left|\mathcal{H}_{mk_n}^n\right.\right)=\mu_r\left(\theta\psi_0\sigma^2_{\frac{m}{M_n}}+\frac{\psi_1\Sigma_U}{\theta}\right)^{\frac{r}{2}}+o_p(1), \tag{A.60}$$

which holds uniformly in m for any even integer $r\geq 2$. Now (A.53) follows from (A.59) and (A.60).

(iii) Following Proposition 2.2.8 in Jacod and Protter (2011), we see that (A.54) boils down to convergence of a Riemann approximation.

This finishes the proof of Theorem 4.1. □

A.3.5 Proof of Theorem 4.2

Proof. We have, by the definition of $\widehat{\mathrm{IV}}_n$ and (A.16) of Lemma A.3 that

$$\begin{aligned}
n^{1/4}\widehat{\mathrm{IV}}_n &= n^{1/4}(\psi_0)^{-1}\left(\sum_{m=0}^{M_n-1}|\overline{Y}_{mk_n}^n|^2-\psi_1\theta^{-2}\widehat{\Sigma}_{U_n}\right),\\
n^{1/4}\mathrm{IV} &= n^{1/4}(\psi_0)^{-1}\left(\frac{1}{\theta M_n}\sum_{m=0}^{M_n-1}\mathbb{E}\left((\beta_m^n)^2\left|\mathcal{H}_{mk_n}^n\right.\right)-\psi_1\theta^{-2}\Sigma_U\right).
\end{aligned}$$

Subtraction gives, due to (A.41) of Lemma A.5 and because $\theta M_n=\sqrt{n}$, that $n^{1/4}(\widehat{\mathrm{IV}}_n-\mathrm{IV})$ equals

$$(\psi_0)^{-1}L_n+Cn^{1/4}(\Sigma_U-\widehat{\Sigma}_{U_n})+o_p(n^{-1/4}),$$

with L_n as defined in (A.33) of Lemma A.4. The first statement of Theorem 4.2 now follows from that Lemma and (A.51), while the second statement is implied by the consistency result in (21). □

A.4 Proofs of the Results in Subsection 4.2

In this subsection we first establish several lemmas to facilitate the proofs of our results in Subsection 4.2. We follow the classical approach in Jacod et al. (2009) and also use several estimates that have been derived in Jacod et al. (2019). Our proofs are often less involved than those in the last paper. This is partly due to our Lemma A.1, which we proved under relatively mild assumptions and from which the higher order moments of the pre-averaged noise process can be easily obtained. Moreover, our setting is not as general as in Jacod et al. (2019).

A.4.1 Auxiliary Lemmas for Subsection 4.2

In the following Lemmas A.6 to A.10, we assume the conditions of Theorem 4.3 are satisfied.

Lemma A.6. *For any $q \geq 1$, we have*

$$\left|\mathbb{E}\left(\widehat{X}_i^n \,|\mathcal{F}_i^n\right)\right| \leq C\Delta_n; \quad \mathbb{E}\left(\left|\widehat{X}_i^n\right|^q |\mathcal{F}_i^n\right) \leq C_q\Delta_n^{q/2}. \tag{A.61}$$

Proof. Using the decomposition in the proof of Lemma 5.2 in Jacod et al. (2009), we have by Itô's formula that $\frac{\widehat{X}_i^n}{2} = \sum_{\ell=1}^3 D_{i,\ell}^n$, where

$$D_{i,1}^n = \int_{i\Delta_n}^{(i+k_n-1)\Delta_n} X_i^n(t)\mathrm{d}M_i^n(t), \quad D_{i,2}^n = b_{i\Delta_n}\int_{i\Delta_n}^{(i+k_n-1)\Delta_n} M_i^n(t)G_i^n(t)\mathrm{d}t,$$
$$D_{i,3}^n = \int_{i\Delta_n}^{(i+k_n-1)\Delta_n} M_i^n(t)(b_t - b_{i\Delta_n})G_i^n(t)\mathrm{d}t + \int_{i\Delta_n}^{(i+k_n-1)\Delta_n} B_i^n(t)\mathrm{d}B_i^n(t).$$

The boundedness of b, σ and g imply that we have that $\mathbb{E}\left(|M_i^n(t)|^q\,|\mathcal{F}_i^n\right) \leq C(k_n\Delta_n)^{q/2}$ and that $\mathbb{E}\left(|B_i^n(t)|^q\,|\mathcal{F}_i^n\right) \leq C(k_n\Delta_n)^q$, and since $k_n = \theta\Delta_n^{-1/2} + o\left(\Delta_n^{-1/4}\right)$ this gives

$$\mathbb{E}\left(\left|D_{i,2}^n\right|^q |\mathcal{F}_i^n\right) \leq C_q\Delta_n^{3q/4}; \tag{A.62}$$
$$\mathbb{E}\left(\left|D_{i,3}^n\right|^q |\mathcal{F}_i^n\right) \leq C_q\Delta_n^{q}. \tag{A.63}$$

The boundedness of σ and g also establish that $|\mathbb{E}\left(M_i^n(t)\,|\mathcal{F}_i^n\right)| = 0$ which gives, together with the boundedness of b and g, that

$$\left|\mathbb{E}\left(D_{i,2}^n\,|\mathcal{F}_i^n\right)\right| \leq C\Delta_n. \tag{A.64}$$

The martingale property of M yields $\mathbb{E}\left(D_{i,1}^n\,|\mathcal{F}_i^n\right) = \mathbb{E}\left(\int_{i\Delta_n}^{(i+k_n-1)\Delta_n} X_i^n(t)\mathrm{d}M_i^n(t)\,|\mathcal{F}_i^n\right) = 0$ and combining this with (A.62) and (A.64) proves the first part of (A.61). The second part of (A.61) follows from (A.62), (A.63) and

$$\mathbb{E}\left(\left|D_{i,1}^n\right|^q |\mathcal{F}_i^n\right) \leq C_q\Delta_n^{q/2},$$

which can be obtained by applying the Burkholder-Davis-Gundy inequalities. This finishes the proof. □

Lemma A.7. *For any $p \geq 2$, we have*

$$\mathbb{E}\Big(\mathbb{E}\left((\zeta(p)_i^n)^4\,|\mathcal{K}_i^n\right)\Big) \leq C_p; \tag{A.65}$$
$$\mathbb{E}\Big(\left(\mathbb{E}\left(\zeta(p)_i^n\,|\mathcal{K}_i^n\right)\right)^2\Big) \leq C_p\Delta_n. \tag{A.66}$$

Proof. We have by Lemma A.1 that

$$\mathbb{E}\left(\left(\overline{U}_i^n\right)^8\right) \leq C\Delta_n^2, \quad \mathbb{E}\left(\left(\overline{U}_i^n\right)^4\right) \leq C\Delta_n, \quad \alpha_n = \mathbb{E}\left(\left(\overline{U}_i^n\right)^2\right) \leq C\Delta_n^{1/2}. \tag{A.67}$$

This implies

$$\mathbb{E}\left(\mathbb{E}\left(\left(\widehat{U}_i^n\right)^4 |\mathcal{K}_i^n\right)\right) \le C\left(\mathbb{E}\left(\left(\overline{U}_i^n\right)^8\right) + \alpha_n^4\right) \le C\Delta_n^2. \tag{A.68}$$

Hölder's inequality gives $\left(\left(\zeta(p)_j^n\right)^4\right) \le C_p k_n^3 \sum_{j=i}^{i+pk_n-1}\left(\left(\widehat{X}_i^n\right)^4 + \left(\widehat{U}_i^n\right)^4 + 2\left(\overline{U}_i^n\right)^4\left(\overline{X}_i^n\right)^4\right)$. Now (A.67), (A.68), together with the second part of (A.61), the independence of X and U, and (A.6) yield (A.65).

We now turn to (A.66). By (A.4) we have

$$\sum_{j=i}^{i+pk_n-1}\mathbb{E}\left(\left(\mathbb{E}\left(\widehat{U}_j^n \middle| \mathcal{G}^n_{i-\left[\frac{k_n}{2}\right]}\right)\right)^2\right) \le \sum_{j=i}^{i+pk_n-1}\frac{C}{(j-i+[k_n/2])^{2v}} \le \frac{C}{k_n^{2v-1}} \le C(\Delta_n)^{v-\frac{1}{2}}. \tag{A.69}$$

On the other hand, for any $i \le j \le i+pk_n-1$, we have (since $\overline{g}_k^n \le C/k_n$)

$$\mathbb{E}\left(\left(\mathbb{E}\left(\overline{U}_j^n \middle| \mathcal{G}^n_{i-\left[\frac{k_n}{2}\right]}\right)\right)^2\right) \le \frac{C}{k_n}\sum_{k=0}^{k_n-1}\mathbb{E}\left(\left(\mathbb{E}\left(U_{j+k}^n \middle| \mathcal{G}^n_{i-\left[\frac{k_n}{2}\right]}\right)\right)^2\right) \le \frac{C}{k_n\left(j-i+\left[\frac{k_n}{2}\right]\right)^{2v-1}},$$

whence

$$\sum_{j=i}^{i+pk_n-1}\mathbb{E}\left(\left(\mathbb{E}\left(\overline{U}_j^n \middle| \mathcal{G}^n_{i-\left[\frac{k_n}{2}\right]}\right)\right)^2\right) \le \frac{C}{k_n^{2v-1}} \le C(\Delta_n)^{v-1/2}. \tag{A.70}$$

By the independence of X and U, (A.6) and (A.61), we deduce

$$\begin{aligned}\mathbb{E}\left(\zeta(p)_i^n\,|\mathcal{K}_i^n\right) &= \sum_{j=i}^{i+pk_n-1}\mathbb{E}\left(\widehat{X}_i^n + \widehat{U}_i^n + 2\overline{U}_i^n\overline{X}_i^n\,|\mathcal{K}_i^n\right) \\ &\le C_p\Delta_n^{1/2} + \sum_{j=i}^{i+pk_n-1}\mathbb{E}\left(\widehat{U}_j^n \middle| \mathcal{G}^n_{i-\left[\frac{k_n}{2}\right]}\right) + 2\Delta_n^{1/2}\sum_{j=i}^{i+pk_n-1}\mathbb{E}\left(\overline{U}_j^n \middle| \mathcal{G}^n_{i-\left[\frac{k_n}{2}\right]}\right).\end{aligned} \tag{A.71}$$

Since $v > 2$, we can now apply Hölder's inequality to the square of this expression to get (A.66) from (A.69), (A.70) and (A.71). $\square$

Lemma A.8. *Let $t_{j,p}^n = j(p+1)k_n\Delta_n$ and define $\Xi_{ij} = -\int_0^1 s\phi_i(s)\phi_j(s)\mathrm{d}s$, and $\Lambda_{ij}(p) = p\Phi_{ij} + \Xi_{ij}$ for $i,j \in \{0,1\}$. We then have*

$$\mathbb{E}\left(\left|\mathbb{E}\left(\left(\eta(p)_j^n\right)^2 - \frac{4k_n^2\Delta_n^2\sigma^4_{t^n_{j,p}}}{\psi_0^2}\Lambda_{00}(p) - \frac{4\Delta_n\Sigma_U^2}{\theta^2\psi_0^2}\Lambda_{11}(p) - \frac{8\Delta_n\sigma^2_{t^n_{j,p}}\Sigma_U}{\psi_0^2}\Lambda_{01}(p)\,|\mathcal{J}(p)_j^n\right)\right|\right) \le C_p\Delta_n^{5/4}. \tag{A.72}$$

Proof. First, we note that

$$(\zeta(p)_i^n)^2 = \sum_{j,j'=0}^{pk_n-1} (\widehat{X}_{i+j}^n \widehat{X}_{i+j'}^n + \widehat{U}_{i+j}^n \widehat{U}_{i+j'}^n + \widehat{X}_{i+j}^n \widehat{U}_{i+j'}^n + \widehat{U}_{i+j}^n \widehat{X}_{i+j'}^n + 4\overline{X}_{i+j}^n \overline{U}_{i+j}^n \overline{X}_{i+j'}^n \overline{U}_{i+j'}^n$$
$$+ 2\widehat{X}_{i+j}^n \overline{X}_{i+j'}^n \overline{U}_{i+j'}^n + 2\overline{X}_{i+j}^n \overline{U}_{i+j}^n \widehat{X}_{i+j'}^n + 2\widehat{U}_{i+j}^n \overline{X}_{i+j'}^n \overline{U}_{i+j'}^n + 2\overline{X}_{i+j}^n \overline{U}_{i+j}^n \widehat{U}_{i+j'}^n).$$

Applying the estimate (A.20) in Jacod et al. (2019), we get

$$\mathbb{E}\left(\left|\sum_{j,j'=0}^{pk_n-1} \mathbb{E}\left(\widehat{X}_{i+j}^n \widehat{X}_{i+j'}^n - 4(\sigma_i^n)^4 \sum_{j,j'=0}^{pk_n-1} \overline{G}_i^n(j,j') \,|\mathcal{F}_i^n\right)\right|\right) \leq C_p \Delta_n^{1/4}. \tag{A.73}$$

Another estimate gives (see the proof of Lemma A.5 in Jacod et al. (2019)):

$$\left|\mathbb{E}\left(\overline{X}_{i+j}^n \overline{X}_{i+j'}^n - (\sigma_i^n)^2 \widehat{G}_i^n(j,j') \,|\mathcal{F}_i^n\right)\right| \leq C\Delta_n^{3/4},$$

which yields, since X and U are independent while $\mathbb{E}\left(\overline{U}_{i+j}^n \overline{U}_{i+j'}^n\right) \leq C\Delta_n^{1/2}$, that

$$\mathbb{E}\left(\left|\sum_{j,j'=0}^{pk_n-1} \mathbb{E}\left(\overline{X}_{i+j}^n \overline{U}_{i+j}^n \overline{X}_{i+j'}^n \overline{U}_{i+j'}^n - (\sigma_i^n)^2 \widehat{G}_i^n(j,j') \mathbb{E}\left(\overline{U}_{i+j}^n \overline{U}_{i+j'}^n\right) |\mathcal{K}_i^n\right)\right|\right) \leq C_p \Delta_n^{1/4}. \tag{A.74}$$

A direct application of (A.4) then leads to

$$\mathbb{E}\left(\sum_{j,j'=0}^{pk_n-1} \left|\mathbb{E}\left(\widehat{U}_{i+j}^n \widehat{U}_{i+j'}^n \,\middle|\mathcal{G}_{i-[\frac{k_n}{2}]}^n\right) - \mathbb{E}\left(\widehat{U}_{i+j}^n \widehat{U}_{i+j'}^n\right)\right|\right) \leq C_p k_n^{-(v-2)} \leq C_p \Delta_n^{1/4}, \tag{A.75}$$

since $v > 5/2$.

We now find bounds for the six remaining terms in the conditional expectation of $(\zeta(p)_i^n)^2$ using symmetry. We first apply the JS-Lemma (A.5) to derive

$$\mathbb{E}\left(\mathbb{E}\left(\widehat{U}_{i+j}^n \,\middle|\mathcal{G}_{i-[\frac{k_n}{2}]}\right)^2\right) \leq C\mathbb{E}\left(\left(\widehat{U}_{i+j}^n\right)^2\right)(j+k_n/2)^{-2v} \leq C\Delta_n (j+k_n/2)^{-2v}, \tag{A.76}$$
$$\mathbb{E}\left(\mathbb{E}\left(\overline{U}_{i+j}^n \,\middle|\mathcal{G}_{i-[\frac{k_n}{2}]}\right)^2\right) \leq C\mathbb{E}\left(\left(\overline{U}_{i+j}^n\right)^2\right)(j+k_n/2)^{-2v} \leq C\sqrt{\Delta_n}(j+k_n/2)^{-2v}.$$

We use this to find, by the independence of X and U and using (A.61), that

$$\mathbb{E}\left(\left|\mathbb{E}\left(\sum_{j,j'=0}^{pk_n-1} \widehat{X}_{i+j}^n \widehat{U}_{i+j'}^n \,|\mathcal{K}_i^n\right)\right|\right) \leq Cpk_n\Delta_n \sum_{j=0}^{pk_n-1} \mathbb{E}\left(\left|\mathbb{E}\left(\widehat{U}_{i+j}^n \,\middle|\mathcal{G}_{i-[\frac{k_n}{2}]}\right)\right|\right)$$
$$\leq C_p\sqrt{\Delta_n} \sum_{j=0}^{pk_n-1} \sqrt{\Delta_n (j+k_n/2)^{-2v}} \leq C_p \Delta_n^{(v+1)/2}. \tag{A.77}$$

For a second estimate, we apply the Cauchy-Schwarz inequality, the independence of X and U, and the bounds of (A.6) and (A.61); this shows

$$\mathbb{E}\left(\left|\mathbb{E}\left(\sum_{j,j'=0}^{pk_n-1}\widehat{X}^n_{i+j}\overline{X}^n_{i+j'}\overline{U}^n_{i+j'}\,|\mathcal{K}^n_i\right)\right|\right)\leq C\Delta_n^{1/2}\Delta_n^{1/4}pk_n\sum_{j'=0}^{pk_n-1}\mathbb{E}\Big(\Big|\mathbb{E}\left(\overline{U}^n_{i+j'}\,\Big|\mathcal{G}_{i-[\frac{k_n}{2}]}\right)\Big|\Big)$$

$$\leq C_p\Delta_n^{1/4}\sum_{j'=0}^{pk_n-1}\sqrt{\sqrt{\Delta_n}(j'+k_n/2)^{-2v}}\leq C\Delta_n^{v/2}. \tag{A.78}$$

For a third estimate, we use that we know from (A.67) and (A.68) that $n^{1/2}\widehat{U}^n_{i+j}$ and $n^{1/4}\widehat{U}^n_{i+j}$ are sequences of stochastic variables with variances that converge to one. Together with the estimates in (A.6), this gives

$$\mathbb{E}\left(\left|\mathbb{E}\left(\sum_{j,j'=0}^{pk_n-1}\widehat{U}^n_{i+j}\overline{X}^n_{i+j'}\overline{U}^n_{i+j'}\,|\mathcal{K}^n_i\right)\right|\right)\leq C\Delta_n^{1/2}\sum_{j,j'=0}^{pk_n-1}\Delta_n^{3/4}\mathbb{E}\left(\left|\mathbb{E}\left(\frac{\overline{U}^n_{i+j'}}{\Delta_n^{1/4}}\frac{\widehat{U}^n_{i+j}}{\Delta_n^{1/2}}\,\Big|\mathcal{G}_{i-[\frac{k_n}{2}]}\right)\right|\right)$$

$$\leq C_p\Delta_n^{5/4}k_n\leq C_p\Delta_n^{3/4}. \tag{A.79}$$

Lemmas A.9 and A.10 in Jacod et al. (2019) yield

$$\begin{aligned}\left|\sum_{j,j'=0}^{pk_n-1}\overline{G}^n_i(j,j')-k_n^4\Delta_n^2\Lambda_{00}(p)\right|&\leq C_p\Delta_n^{1/2};\\ \left|\sum_{j,j'=0}^{pk_n-1}\widehat{G}^n_i(j,j')\mathbb{E}\Big(\overline{U}^n_{i+j}\overline{U}^n_{i+j'}\Big)-2k_n^2\Delta_n\Lambda_{01}(p)\Sigma_U\right|&\leq C_p\Delta_n^{1/2};\\ \left|\sum_{j,j'=0}^{pk_n-1}\mathbb{E}\Big(\widehat{U}^n_{i+j}\widehat{U}^n_{i+j'}\Big)-4\Lambda_{11}(p)\Sigma_U^2\right|&\leq C_p\Delta_n^{1/2}.\end{aligned} \tag{A.80}$$

Now the result follows from (A.73)-(A.80). □

Lemma A.9. *For any $p\geq 2$, we have*

$$\Delta_n^{-1/4}F(p)_n\xrightarrow{\mathbb{P}}0; \tag{A.81}$$

$$\Delta_n^{-1/4}F'(p)_n\xrightarrow{\mathbb{P}}0; \tag{A.82}$$

$$\Delta_n^{-1/4}\widehat{C}(p)_n\xrightarrow{\mathbb{P}}0; \tag{A.83}$$

$$\Delta_n^{-1/4}\widehat{C}'(p)_n\xrightarrow{\mathbb{P}}0; \tag{A.84}$$

$$\Delta_n^{-1/4}\widehat{C}''_n\xrightarrow{\mathbb{P}}0; \tag{A.85}$$

$$\mathbb{E}\left(\sup_{t\leq T}\left(M'(p)_n\right)^2\right)\leq C_p\sqrt{\Delta_n}. \tag{A.86}$$

Proof. We prove these equations in a number of separate steps.

(1) Proof of (A.81) and (A.82). First, we note that due to (A.76) we have $\mathbb{E}\Big(\Big|\mathbb{E}\left(\widehat{U}^n_i\,|\mathcal{K}^n_i\right)\Big|\Big)\leq Ck_n^{-v}$.

Together with (A.61), and the independence of X and U, we get

$$\mathbb{E}(|\mathbb{E}\left(\Psi_i^n\,|\mathcal{K}_i^n\right)|) \leq C\Delta_n, \quad \mathbb{E}(|\mathbb{E}\left(\zeta(p)_i^n\,|\mathcal{K}_i^n\right)|) \leq Cpk_n\Delta_n.$$

Since $K_n^p \leq \frac{C}{pk_n\Delta_n}$, we have $\mathbb{E}(|F(p)_n|) \leq C\Delta_n^{\frac{1}{2}}$. The same result holds for $F'(p)_n$. Now (A.81) and (A.82) follow.

(2) Proof of (A.83). From the estimates (A.61), (A.6) and (A.67), we have

$$\mathbb{E}\Big((\Psi_i^n)^2\Big) \leq C\Delta_n. \tag{A.87}$$

Since $n - k_n - I_n^p \leq C_p/\sqrt{\Delta_n}$, the claim follows.

(3) Proof of (A.84). Let $\Gamma_n = \sum_{|\ell|\leq \ell_n} \gamma(\ell)$. Then (18) implies

$$\alpha_n - \frac{T\psi_1\Gamma_n}{\theta\sqrt{\Delta_n}(n-k_n+2)} = \alpha_n - \frac{\psi_1\Gamma_n}{k_n} + o(\Delta_n^{\frac{3}{4}}). \tag{A.88}$$

Since $\alpha_n = \frac{1}{k_n}\sum_{|\ell|\leq k_n-1} \phi_1^n(\ell)\gamma(\ell)$, we have

$$\begin{aligned}
\left|\alpha_n - \frac{\psi_1\Gamma_n}{k_n}\right| &= \frac{1}{k_n}\left|\sum_{|\ell|\leq k_n-1} \phi_1^n(\ell)\gamma(\ell) - \psi_1 \sum_{|\ell|\leq \ell_n}\gamma(\ell)\right| \\
&\leq \frac{1}{k_n}\left|\psi_1 \sum_{\ell_n<|\ell|\leq k_n-1}\gamma(\ell)\right| + \frac{1}{k_n}\left|\sum_{|\ell|\leq k_n-1}\gamma(\ell)\left(\phi_1^n(\ell)-\psi_1\right)\right| \\
&\leq \frac{C}{k_n\ell_n^{v-1}} + \frac{C}{k_n}\sum_{|\ell|\leq k_n-1}\frac{\gamma(\ell)\ell}{k_n} \leq C\Delta_n^{1\wedge\left(\frac{1}{2}+(v-1)\kappa\right)},
\end{aligned} \tag{A.89}$$

where the second inequality is due to (A.10) and the last inequality follows from the fact that $v>2$ so that $\sum\gamma(\ell)\ell<\infty$, while $\ell_n \asymp \Delta_n^{-\kappa}$. Then (A.88) and (A.89) imply

$$\left|\frac{(n-k_n+2)\,\alpha_n\sqrt{\Delta_n}}{\theta\psi_0} - \frac{\psi_1}{\theta^2\psi_0}\Gamma_n\right| \leq C\Delta_n^{\frac{1}{2}\wedge(v-1)\kappa} + o(\Delta_n^{1/4}).$$

Since $(v-1)\kappa > 1/4$ we have

$$\Delta_n^{-1/4}\left(\frac{(n-k_n+2)\,\alpha_n\sqrt{\Delta_n}}{\theta\psi_0} - \frac{\psi_1}{\theta^2\psi_0}\Gamma_n\right) \to 0. \tag{A.90}$$

On the other hand, we have by (A.51) that

$$\Delta_n^{-1/4}\left(\frac{\psi_1}{\theta^2\psi_0}\left(\Gamma_n - \sum_{|\ell|\leq\ell_n}\widehat{\gamma(\ell)}_n\right)\right) \xrightarrow{\mathbb{P}} 0. \tag{A.91}$$

Now (A.84) is proven by (A.90) and (A.91).

(4) Proof of (A.85): see Lemma 5.5 of Jacod et al. (2009).

(5) For (A.86), we apply Hölder's inequality and (A.87) to conclude that $\mathbb{E}\Big(\big(\eta'(p)_j^n\big)^2\Big) \leq C_p\Delta_n$. Doob's inequality and the fact that $K_n^p \leq C_p/\sqrt{\Delta_n}$ then together imply

$$\mathbb{E}\Big(\sup_{t\leq T}\left(M'(p)_n\right)^2\Big) \leq 4\sum_{j=0}^{K_n^p}\mathbb{E}\Big(\big(\eta'(p)_j^n\big)^2\Big) \leq C_p\sqrt{\Delta_n}.$$

$\square$

Lemma A.10. *For any $p \geq 2$, the sequence $\Delta_n^{-1/4}M(p)_n$ of processes converges stably in law to*

$$\Upsilon_1(p) = \int_0^1 V(p)_s \mathrm{d}W'_s, \tag{A.92}$$

where W' is as in Theorem 4.3 and $V(p)_t$ is the square root of

$$V(p)_t^2 = \frac{4}{\psi_0^2(p+1)}\left(\Lambda_{00}(p)\theta\sigma_t^4 + 2\Lambda_{01}(p)\frac{\sigma_t^2\Sigma_U}{\theta} + \Lambda_{11}(p)\frac{\Sigma_U^2}{\theta^3}\right). \tag{A.93}$$

Proof. In view of the classical central limit theorems for triangular arrays of martingale differences in, e.g., Theorem IX.7.28 in Jacod and Shiryaev (2003), it suffices to prove the following:

$$\frac{1}{\sqrt{\Delta_n}}\sum_{j=0}^{K_n^p}\left(\mathbb{E}\left(\left(\eta(p)_j^n\right)^2 \middle| \mathcal{J}(p)_j^n\right) - \left(\overline{\eta}(p)_j^n\right)^2\right) \xrightarrow{\mathbb{P}} \int_0^1 V(p)_s^2\mathrm{d}s, \tag{A.94}$$

$$\frac{1}{\Delta_n}\sum_{j=0}^{K_n^p}\mathbb{E}\left(\left(\eta(p)_j^n\right)^4 \middle| \mathcal{J}(p)_j^n\right) \xrightarrow{\mathbb{P}} 0, \tag{A.95}$$

$$\frac{1}{\Delta_n^{1/4}}\sum_{j=0}^{K_n^p}\mathbb{E}\left(\left(\eta(p)_j^n\right)\Delta(N,p)_j^n \middle| \mathcal{J}(p)_j^n\right) \xrightarrow{\mathbb{P}} 0, \tag{A.96}$$

where $\Delta(Z,p)_j^n = Z_{(j+1)(p+1)k_n\Delta_n} - Z_{j(p+1)k_n\Delta_n}$, and N is any bounded martingale orthogonal to W, or $N = W$.

1. Proof of (A.94). Equation (A.66) implies $\mathbb{E}\Big(\big(\overline{\eta}(p)_j^n\big)^2\Big) \leq C_p\Delta_n^2$, whence $\frac{1}{\sqrt{\Delta_n}}\sum_{j=0}^{K_n^p}\left(\overline{\eta}(p)_j^n\right)^2 \xrightarrow{\mathbb{P}} 0$. The estimate (A.72), plus Riemann integration, and (18) yield

$$\frac{1}{\sqrt{\Delta_n}}\sum_{j=0}^{K_n^p}\mathbb{E}\left(\left(\eta(p)_j^n\right)^2 \middle| \mathcal{J}(p)_j^n\right) \xrightarrow{\mathbb{P}} \int_0^1 V(p)_s^2\mathrm{d}s.$$

2. Proof of (A.95). By (A.65) we have $\Delta_n^{-2}\mathbb{E}\Big(\big(\eta(p)_j^n\big)^4\Big) \leq C$ so the Markov inequality gives $\mathbb{E}\left(\left(\eta(p)_j^n\right)^4 \middle| \mathcal{J}(p)_j^n\right) = O_p(\Delta_n^2)$. Then (A.95) follows since $K_n^p \leq C_p\Delta_n^{-\frac{1}{2}}$.

3. Proof of (A.96). Let $\ell(p)_j^n = j(p+1)k_n, \bar{\ell}(p)_j^n = \ell(p)_j^n - \left[\frac{k_n}{2}\right]$. It is equivalent to prove

$$\Delta_n^{1/4} \sum_{j=0}^{K_n^p} \mathbb{E}\left(\zeta(p)^n_{\ell(p)_j^n} \Delta(N,p)_j^n \,|\mathcal{J}(p)_j^n\right) \xrightarrow{\mathbb{P}} 0.$$

In view of (5.62) of Jacod et al. (2009), it then suffices to prove that

$$L_n(p) := \Delta_n^{1/4} \sum_{j=0}^{K_n^p} \mathbb{E}\left(\sum_{i=\ell(p)_j^n}^{\ell(p)_j^n + pk_n - 1} (\widehat{U}_i^n + 2\overline{X}_i^n \overline{U}_i^n)\Delta(N,p)_j^n \,|\mathcal{J}(p)_j^n\right) \xrightarrow{\mathbb{P}} 0, \tag{A.97}$$

where we may assume that N is a square-integrable martingale.

By the independence of $\mathcal{F}$ and $\mathcal{G}$, we have

$$\sum_{i=\ell(p)_j^n}^{\ell(p)_j^n + pk_n - 1} \left|\mathbb{E}\left(\widehat{U}_i^n \Delta(N,p)_j^n \,|\mathcal{J}(p)_j^n\right)\right| \leq \sqrt{\mathbb{E}\left(\left(\Delta(N,p)_j^n\right)^2 \left|\mathcal{F}^n_{\bar{\ell}(p)_j^n}\right.\right)}\Theta(p)_j^n;$$

$$\sum_{i=\ell(p)_j^n}^{\ell(p)_j^n + pk_n - 1} \left|\mathbb{E}\left(\overline{X}_i^n \overline{U}_i^n \Delta(N,p)_j^n \,|\mathcal{J}(p)_j^n\right)\right| \leq \sqrt{\mathbb{E}\left(\left(\Delta(N,p)_j^n\right)^2 \left|\mathcal{F}^n_{\bar{\ell}(p)_j^n}\right.\right)}\overline{\Theta}(p)_j^n;$$

where

$$\Theta(p)_j^n := \sum_{i=\ell(p)_j^n}^{\ell(p)_j^n + pk_n - 1} \left|\mathbb{E}\left(\widehat{U}_i^n \left|\mathcal{G}^n_{\bar{\ell}(p)_j^n}\right.\right)\right|; \quad \overline{\Theta}(p)_j^n := \sum_{i=\ell(p)_j^n}^{\ell(p)_j^n + pk_n - 1} \left|\mathbb{E}\left(\overline{U}_i^n \left|\mathcal{G}^n_{\bar{\ell}(p)_j^n}\right.\right)\right| \sqrt{\Delta_n^{1/2}}.$$

Note that we have used (A.6) to bound $\mathbb{E}\left(\left(\overline{X}_i^n\right)^2 \left|\mathcal{F}^n_{\bar{\ell}(p)_j^n}\right.\right)$ by $\Delta_n^{1/2}$. We find that

$$L_n(p)^2 \leq \sqrt{\Delta_n}\left(\sum_{j=0}^{K_n^p} \sqrt{\mathbb{E}\left(\left(\Delta(N,p)_j^n\right)^2 \left|\mathcal{F}^n_{\bar{\ell}(p)_j^n}\right.\right)}\left(\Theta(p)_j^n + 2\overline{\Theta}(p)_j^n\right)\right)^2.$$

Repeated applications of the JS-Lemma and the independence of $\mathcal{G}$ and $\mathcal{F}$ give

$$\mathbb{E}\Big(\left(\Theta(p)_j^n\right)^2\Big) \leq \frac{C_p}{k_n^{2(v-1)}}; \quad \mathbb{E}\Big(\left(\overline{\Theta}(p)_j^n\right)^2\Big) \leq \frac{C_p}{k_n^{2v}}, \tag{A.98}$$

so we have

$$\begin{aligned}
\mathbb{E}\Big(\left(L_n(p)\right)^2\Big) &\leq \sqrt{\Delta_n}\mathbb{E}\left(\sum_{j=0}^{K_n^p}\left(\Delta(N,p)_j^n\right)^2\right)\mathbb{E}\left(\sum_{j=0}^{K_n^p}\left(\Theta(p)_j^n + 2\overline{\Theta}(p)_j^n\right)^2\right) \\
&\leq C_p \mathbb{E}\Big(\left(N_1 - N_0\right)^2\Big)\Delta_n^{v-1} \to 0.
\end{aligned}$$

The first inequality is an application of Cauchy-Schwarz inequality, and the second one is due to the fact that N is a square-integrable martingale, the estimate (A.98) and the fact that $K_n^p \leq C_p/\sqrt{\Delta_n}$.

This completes the proof of Lemma A.10. $\square$

A.4.2 Proof of Theorem 4.3

Proof. We invoke Lemmas A.9 and A.10, which in turn rely on Lemmas A.6, A.7 and A.8. Recalling the decomposition in (A.2), we note that we have proved in Lemma A.9 that

$$\lim_{p\to\infty}\limsup_{n\to\infty}\mathbb{P}(|Q(p)_n|\geq\varepsilon)=0,$$

for any $\varepsilon>0$, where

$$Q(p)_n:=\Delta_n^{-1/4}\left(M'(p)_n+F(p)_n+F'(p)_n+\widehat{C}(p)_n+\widehat{C}'(p)_n+\widehat{C}''_n\right).$$

Lemma A.10 shows convergence of $\Delta_n^{-1/4}M(p)_n$ to $\Upsilon_1(p)$, and for the fixed Brownian motion W' we have that $V(p)_t(\omega)$ converges pointwise to $V_t(\omega)$ so $\Upsilon(p)_1\xrightarrow{\mathbb{P}}\Upsilon_1$. This proves Theorem 4.3. $\square$

A.4.3 Proof of Corollary 4.2

Proof. The result $\widehat{\Sigma}_n\xrightarrow{\mathbb{P}}\int_0^1 V_t^2\mathrm{d}t$ follows from the following convergence in probability results:

$$\widehat{\Sigma}_{U_n}\xrightarrow{\mathbb{P}}\Sigma_U,\quad \widetilde{\mathrm{IV}}_n\xrightarrow{\mathbb{P}}\mathrm{IV};\ \sum_{i=0}^{n-k_n+1}\left(\overline{Y}_i^n\right)^4\xrightarrow{\mathbb{P}}\int_0^1\left(3\theta^2\psi_0^2\sigma_t^4+6\psi_0\psi_1\sigma_t^2\Sigma_U+\frac{3}{\theta^2}\psi_1^2\Sigma_U^2\right)\mathrm{d}t.$$

The first two statements follow from Proposition 3.3 and Theorem 4.3, whereas the last one is due to (5.65) in Jacod et al. (2009) when we replace the asymptotic variance of pre-averaged noise (called α_t in that paper) by Σ_U, and this can be done because of our Lemma A.1. $\square$

A.5 Proofs of the Results in Section 5

A.5.1 Proof of Theorem 5.1

Proof. By Theorem 4.3 we have $\widetilde{\mathrm{IV}}_n\xrightarrow{\mathbb{P}}\mathrm{IV}$, by Proposition 3.3 we have $\widehat{\Sigma}_{U_n}\xrightarrow{\mathbb{P}}\Sigma_U$ and by Proposition 3.1 we have $\widehat{\Sigma}_{U_n}^{(1)}=\langle Y,Y\rangle(1)_n\xrightarrow{\mathbb{P}}\gamma(0)-\gamma(1)$. Therefore

$$\widetilde{\mathrm{IV}}_n^{(1)}=\widetilde{\mathrm{IV}}_n+\frac{\psi_1(\widehat{\Sigma}_{U_n}-\widehat{\Sigma}_{U_n}^{(1)})}{\theta^2\psi_0}\xrightarrow{\mathbb{P}}\mathrm{IV}+\frac{\psi_1(\Sigma_U-\gamma(0)+\gamma(1))}{\theta^2\psi_0}.$$

This shows that

$$\widetilde{\gamma(0)}_n^{(2)}-\widehat{\gamma(0)}_n=O_p(j_n\Delta_n),\quad \widetilde{\langle Y,Y\rangle}(j)_n^{(2)}-\widehat{\langle Y,Y\rangle}(j)_n=O_p(\Delta_n j),$$

which gives

$$\widetilde{\gamma(j)}_n^{(2)}-\widehat{\gamma(j)}_n=\left(\widetilde{\gamma(0)}_n^{(2)}-\widetilde{\langle Y,Y\rangle}(j)_n^{(2)}\right)-\left(\widehat{\gamma(0)}_n-\widehat{\langle Y,Y\rangle}(j)_n\right)\ =\ O_p(\Delta_n(j\vee j_n)),$$
$$\widetilde{\Sigma}_{U_n}^{(2)}-\widehat{\Sigma}_{U_n}=\widetilde{\gamma(0)}_n^{(2)}-\widehat{\gamma(0)}_n+2\sum_{j=1}^{\ell_n}\left(\widetilde{\gamma(j)}_n^{(2)}-\widehat{\gamma(j)}\right)\ =\ O_p((\ell_n^2\vee j_n\ell_n)\Delta_n).$$

The asymptotic conditions (14) then imply that $\Delta_n^{-1/4}\left(\widetilde{\Sigma}_{U_n}^{(2)} - \widehat{\Sigma}_{U_n}\right) \xrightarrow{\mathbb{P}} 0$. This proves (42) for $K = 2$ and the consistency $\widetilde{\Sigma}_{\mathrm{IV}_n}^{(2)} \xrightarrow{\mathbb{P}} \int_0^1 V_t^2 \mathrm{d}t$. It also immediately yields $\Delta_n^{-1/4}\left(\widetilde{\mathrm{IV}}_n^{(2)} - \widetilde{\mathrm{IV}}_n\right) \xrightarrow{\mathbb{P}} 0$.

Now assume we have for a certain $k \geq 2$ that

$$\Delta_n^{-1/4}\left(\widetilde{\Sigma}_{U_n}^{(k)} - \widehat{\Sigma}_{U_n}\right) \xrightarrow{\mathbb{P}} 0; \tag{A.99}$$

$$\Delta_n^{-1/4}\left(\widetilde{\mathrm{IV}}_n^{(k)} - \widetilde{\mathrm{IV}}_n\right) \xrightarrow{\mathbb{P}} 0; \tag{A.100}$$

$$\widetilde{\Sigma}_{\mathrm{IV}_n}^{(k)} \xrightarrow{\mathbb{P}} \int_0^1 V_t^2 \mathrm{d}t. \tag{A.101}$$

A direct calculation shows

$$\begin{aligned}
\widetilde{\Sigma}_{U_n}^{(k+1)} - \widetilde{\Sigma}_{U_n}^{(k)} &= -\frac{(2\ell_n + 1) j_n \widetilde{\mathrm{IV}}_n^{(k)}}{2(n - j_n + 1)} + \sum_{j=1}^{\ell_n} \frac{j \widetilde{\mathrm{IV}}_n^{(k)}}{n - j + 1} = O_p((\ell_n^2 \vee j_n \ell_n)\Delta_n); \\
\widetilde{\mathrm{IV}}_n^{(k+1)} - \widetilde{\mathrm{IV}}_n^{(k)} &= \frac{\psi_1\left(\widetilde{\Sigma}_{U_n}^{(k)} - \widetilde{\Sigma}_{U_n}^{(k+1)}\right)}{\theta^2 \psi_0} = O_p((\ell_n^2 \vee j_n \ell_n)\Delta_n).
\end{aligned}$$

Assumption (14) then implies that (A.99) and (A.100) hold for $k + 1$ as well, and (A.101) then follows. This proves Theorem 5.1. $\square$

B Additional Simulation Studies

In this section, we provide additional Monte Carlo simulation results that assess the effects of price discreteness and correlation between X and U. Price discreteness renders dependence between X and U. The results in Section B.1 show that the presence of minimal ticks has relatively little impact on the estimation of the moments of noise and the IV. Furthermore, the results in Section B.2 show that in the situation when X and U are correlated our multi-step estimators still appear to be performing well.

B.1 Price Discreteness

We consider a setting in which the observed price is rounded to 1 cent. The observed logarithmic price is then given by:

$$Y_t^{\mathrm{rd}} = \log\left([100\exp(Y_t)]/100\right), \tag{B.1}$$

where $[x]$ denotes the integer that is closest to x. Now the microstructure noise has two components:

$$U_i^{\mathrm{rd}} = Y_{i\Delta_n}^{\mathrm{rd}} - X_{i\Delta_n} = \underbrace{Y_{i\Delta_n}^{\mathrm{rd}} - Y_{i\Delta_n}}_{\text{error due to discreteness}} + \underbrace{Y_{i\Delta_n} - X_{i\Delta_n}}_{\text{error due to market microstructure}}. \tag{B.2}$$

Figure B.1 compares our two-step estimators of the second moments of U^{rd} to the true values for the model setup of Section 6. The two-step estimators still yield accurate estimates, although there is a small bias. In the estimation of the integrated volatility, we have a bias of 4.47×10^{-5} and a standard deviation of 3.55×10^{-5}; these are relatively small compared to the expected value of the integrated volatility which is 4.44×10^{-4}.

B.2 Correlation between X and U

We also provide simulation evidence on the robustness of our estimators when dependence between X and U is introduced by choosing a fixed correlation $\rho_{\epsilon W}$ between the process ϵ in (44) and the increments of the Brownian motion W. Table B.1 shows the centered means and standard deviations of $\widetilde{\mathrm{IV}}_n^{(2)}$. Results are shown for the cases $\rho_{\epsilon W} = 0$, $\rho_{\epsilon W} = 0.7$ and $\rho_{\epsilon W} = -0.7$, and for three different values of the tuning parameter: $\theta = 0.4$, $\theta = 0.6$, and the value $\theta = \theta^*$ defined in (28). The results show that our estimator is relatively insensitive to the choice of the tuning parameter θ and to the correlation between X and U for this model specification.

In a second simulation experiment, we investigate the performance of our two-step estimators for the second moments of noise when the increments of the Brownian motion W and the noise component e in (44) are correlated. The fixed correlation coefficient ρ_{eW} was taken to be either 1 or -1. The results in Figure B.2 show that the biases in the estimates are very small, both for a fixed value of θ and for the optimized value θ^*.

ι	-0.7		-0.3		0		0.3		0.7	
$\theta = 0.4$										
$\rho_{\epsilon W} = 0$	-1.33	(3.72)	-0.96	(3.71)	-0.62	(3.71)	-0.14	(3.72)	1.00	(3.78)
$\rho_{\epsilon W} = 0.7$	-1.39	(3.71)	-1.01	(3.72)	-0.60	(3.72)	0.04	(3.74)	2.00	(3.88)
$\rho_{\epsilon W} = -0.7$	-1.26	(3.71)	-0.90	(3.70)	-0.62	(3.70)	-0.31	(3.71)	0.02	(3.73)
$\theta = 0.6$										
$\rho_{\epsilon W} = 0$	-1.00	(4.33)	-0.93	(4.33)	-0.86	(4.33)	-0.77	(4.34)	-0.56	(4.38)
$\rho_{\epsilon W} = 0.7$	-1.00	(4.33)	-0.93	(4.33)	-0.85	(4.34)	-0.72	(4.35)	-0.35	(4.43)
$\rho_{\epsilon W} = -0.7$	-0.99	(4.33)	-0.92	(4.33)	-0.87	(4.33)	-0.81	(4.33)	-0.76	(4.34)
$\theta = \theta^*$										
$\rho_{\epsilon W} = 0$	-1.18	(3.87)	-0.97	(3.91)	-0.80	(3.91)	-0.59	(3.96)	-0.24	(4.13)
$\rho_{\epsilon W} = 0.7$	-1.21	(3.88)	-0.99	(3.90)	-0.79	(3.92)	-0.48	(4.00)	0.00	(4.28)
$\rho_{\epsilon W} = -0.7$	-1.15	(3.90)	-0.96	(3.89)	-0.82	(3.90)	-0.67	(3.92)	-0.59	(3.95)

Table B.1: Estimation of the IV using $\widetilde{\mathrm{IV}}_n^{(2)}$ when X and U are correlated. The numbers represent the centered means with standard deviations between parentheses, based on 1,000 simulations for each scenario. All numbers in the table are multiplied by 10^5. The time step is $\Delta_n = 1$ sec and the number of observations n is 23,400. For the tuning parameters we took $j_n = 20$ and $\ell_n = 10$ while the value of θ varies, as shown in the first column of the table.

C Empirical Study of Transaction Data for General Electric

We collect 2,721,475 transaction prices of General Electric (GE) over the month January 2011. On average there are 5.8 observations per second. In contrast to the analysis of Citigroup transaction prices in Sections 7.2 and 7.3, bias correction plays a very pronounced role here. Despite the high data frequency, the finite sample bias can be very significant if the underlying noise-to-signal ratio is small (recall Remark 3.3). This is indeed the case as Figure C.1 reveals: compared with Citigroup, the data frequency of the General Electric sample is typically lower but the noise-to-signal ratio is also (much) smaller. While the data frequency is immediately available, the noise-to-signal ratio is latent. Therefore, one should always be wary to rely solely on asymptotic theory in practice.

The top panel of Figure C.2 shows that both the realized volatility (RV) and local averaging (LA) estimators indicate that the noise is strongly autocorrelated, while the bias corrected realized volatility (BCRV) estimator reveals that the noise is only weakly dependent. Such a pattern also appears in our simulation study, where we have seen that it is the finite sample bias that induces this discrepancy. Since the dependence in noise is quite weak, we would expect the estimators $\widetilde{\mathrm{IV}}_n^{(1)}$ and $\widetilde{\mathrm{IV}}_n^{(2)}$ to be close to each other, if the latter is accurate. This is indeed the case, as the bottom panel of Figure C.2 shows. However, the other two estimators $\widetilde{\mathrm{IV}}_n$ and $\widetilde{\mathrm{IV}}_n^{\mathrm{JLZ}}$, which don't apply finite sample bias corrections, seem to be biased downwards.

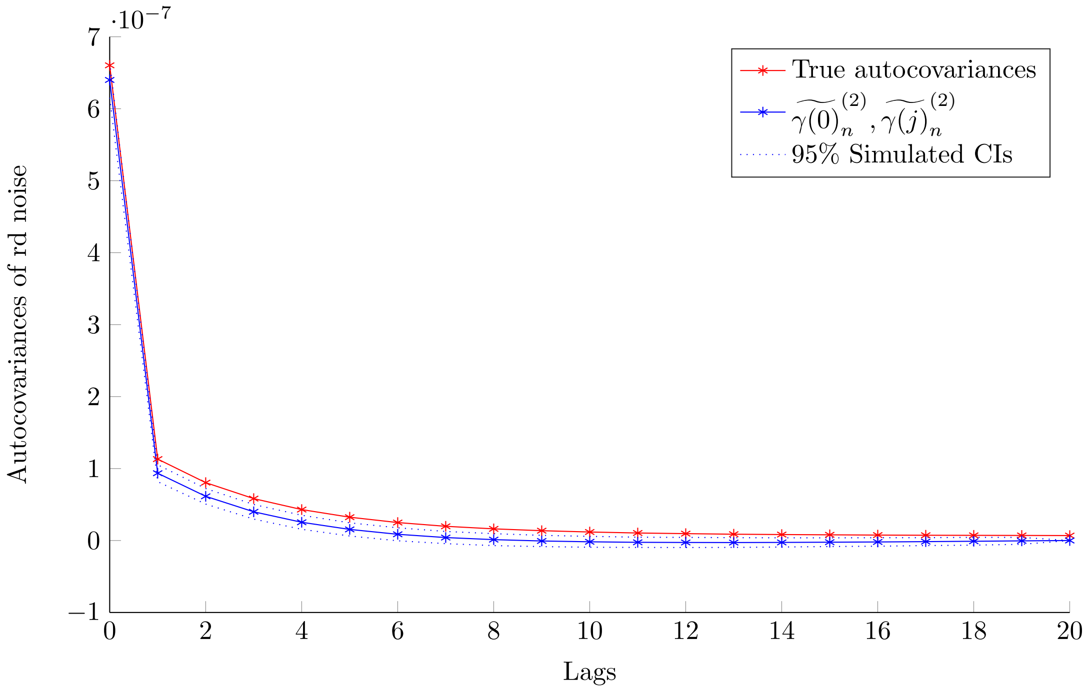

Figure B.1: Estimation of the autocovariances of microstructure noise with rounded prices as specified in (B.1) and (B.2) for the model setup of Section 6. The estimators $\widetilde{\gamma(0)}_n^{(2)}, \widetilde{\gamma(j)}_n^{(2)}$ are defined in (37) and (38). The AR(1)-coefficient of U equals $\iota = 0.7$. The number of simulations is 1,000 and the time step is $\Delta_n = 0.2$ sec. The tuning parameters are $j_n = 20$ and $\ell_n = 10$ and θ is selected according to (28). The "true autocovariances" were determined as the means of the 1,000 sample autocovariances of U^{rd}.

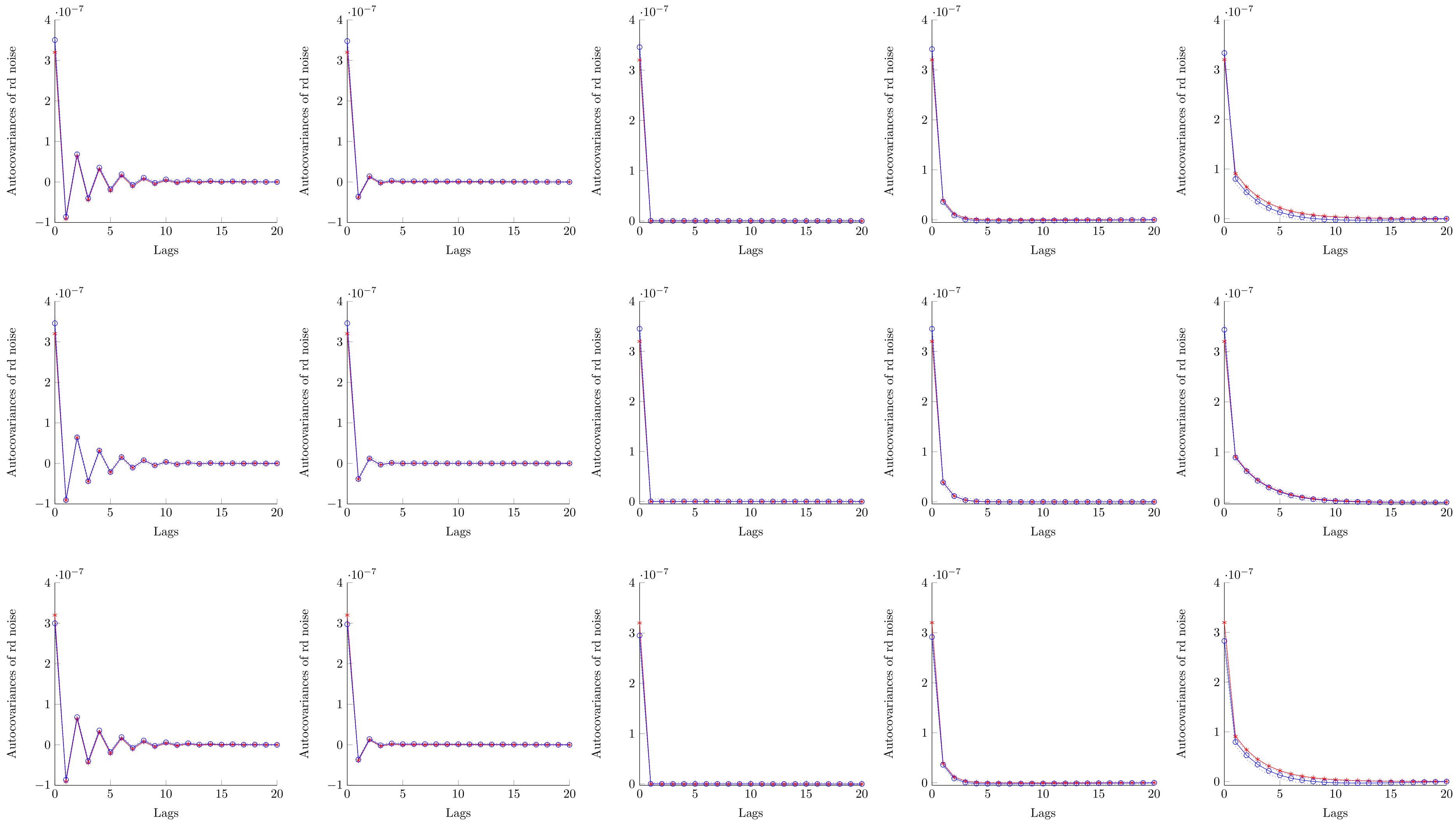

Figure B.2: Estimation of the autocovariances of noise using $\widetilde{\gamma(0)}_n^{(2)}, \widetilde{\gamma(j)}_n^{(2)}$ when X and U are correlated. The red stars are the true values of the autocovariances of noise. The blue circles are the mean estimates of our two-step estimators and the dashed lines are the 95% simulated confidence intervals. The number of simulations is 1,000. $\Delta = 0.2$ sec and the number of observations is 468,000. The tuning parameter of the RV estimator is $j_n = 20$ and $\ell_n = 10$. The three panels, from top to bottom correspond to the specifications $\rho_{eW} = 1$, $\theta = \theta^*$, $\rho_{eW} = 1$, $\theta = 0.6$ and $\rho_{eW} = -1$, $\theta = \theta^*$. The correlation parameter ι varies from left to right in the panels; we show results for ι equal to -0.7, -0.3, 0, 0.3 and 0.7.

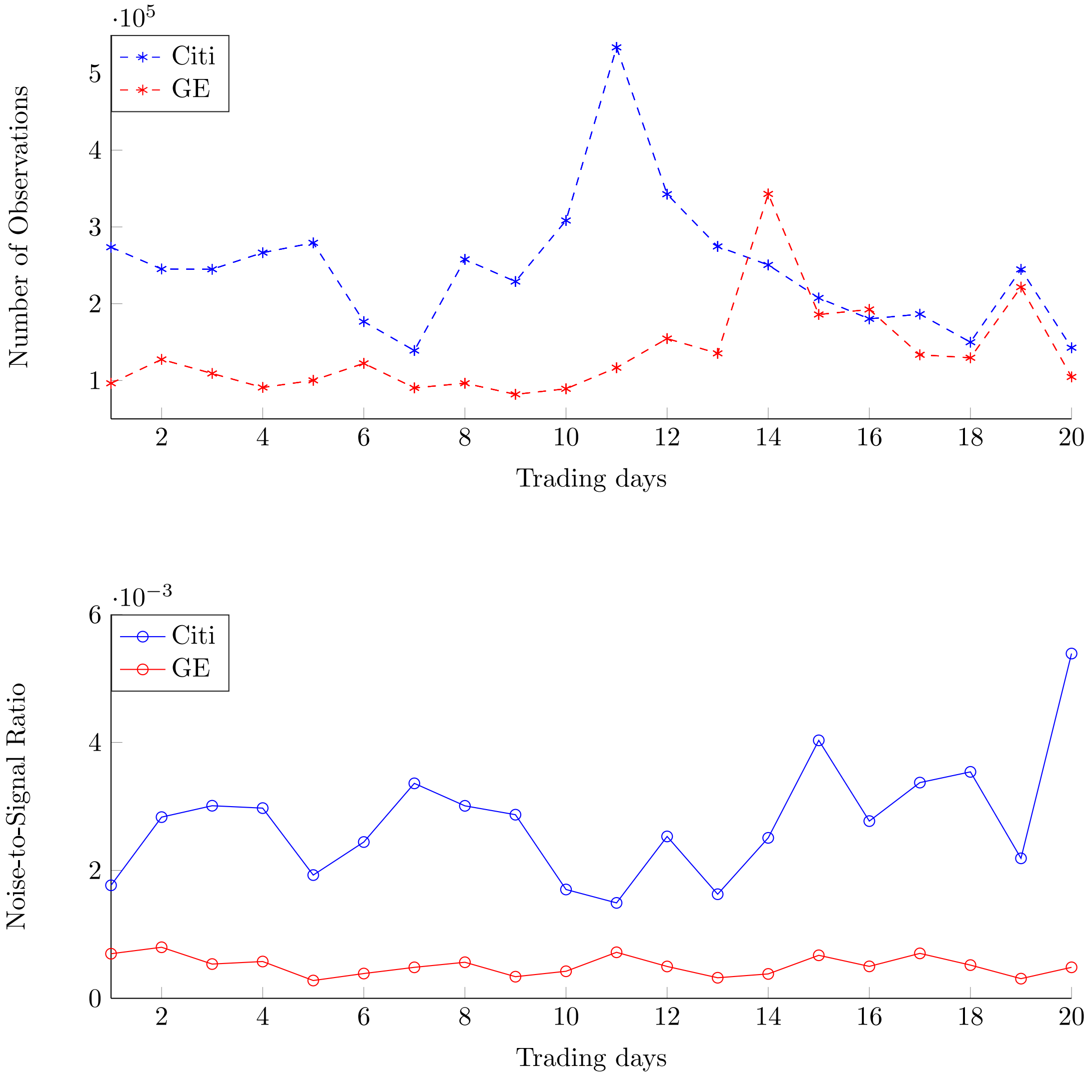

Figure C.1: Number of daily observations of transaction prices (top panel) and noise-to-signal ratio (bottom panel) for Citigroup (C) and General Electric (GE). Sample period: January, 2011, consisting of 20 trading days. In the bottom panel, the noise-to-signal ratio, $\frac{\Sigma_U^2}{\int_0^1 \sigma_s^2 \mathrm{d}s}$, is estimated by $\frac{\widetilde{\Sigma}_{U_n}^{(2)}}{\widetilde{\mathrm{IV}}_n^{(2)}}$, where $\widetilde{\Sigma}_{U_n}^{(2)}$ and $\widetilde{\mathrm{IV}}_n^{(2)}$ are defined in (39) and (40), respectively. We set $j_n = 30$, $\ell_n = 10$ and θ is selected according to (28).

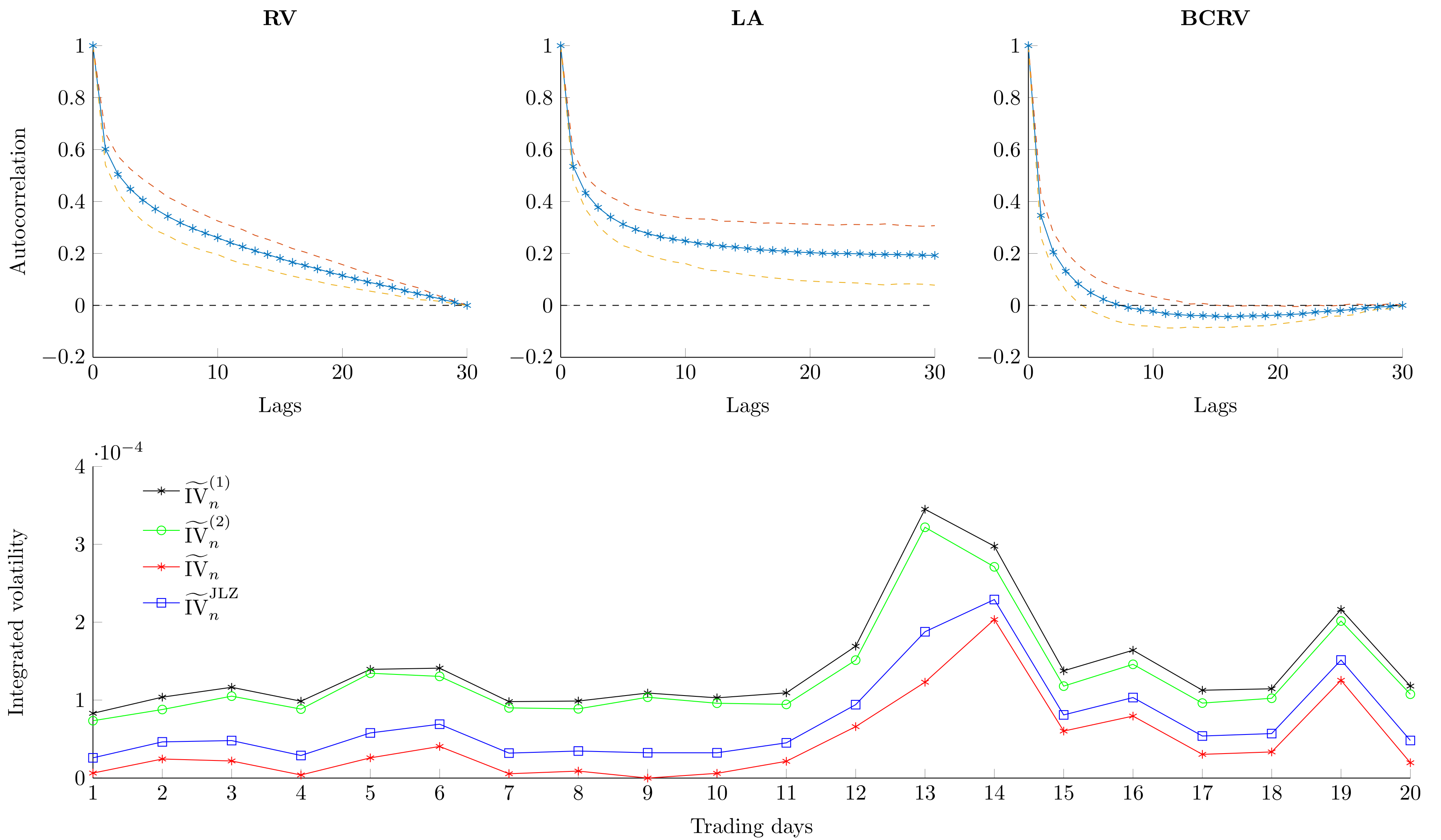

Figure C.2: Autocorrelations of noise and integrated volatility based on transaction data for General Electric (GE). Sample period: January, 2011, consisting of 20 trading days. On average there are 5.8 observations per second in the sample. Top panel: From the left to the right, we display the realized volatility (RV), local averaging (LA), and the bias corrected realized volatility (BCRV) estimators of the autocorrelations of noise against the number of lags j. The three estimators are applied to and then averaged over each of the 20 trading days. The stars indicate the means of the 20 estimates. The dashed lines are 2 standard deviations away from the mean. Bottom panel: Estimation of the integrated volatility. The estimators $\widetilde{\mathrm{IV}}_n^{(1)}$, $\widetilde{\mathrm{IV}}_n^{(2)}$, and $\widetilde{\mathrm{IV}}_n$ are given by (35), (40), and (25). The $\widetilde{\mathrm{IV}}_n^{\mathrm{JLZ}}$ estimator is proposed in Jacod et al. (2019). We set $j_n = 30$, $\ell_n = 10$ and θ is selected according to (28).

References

Barndorff-Nielsen, O. E., S. E. Graversen, J. Jacod, M. Podolskij, and N. Shephard (2006): "A central limit theorem for realised power and bipower variations of continuous semimartingales," in *From Stochastic Calculus to Mathematical Finance*, ed. by Y. Kabanov, R. Lipster, and J. Stoyanov, Springer.

Chung, K. L. (2001): *A Course in Probability Theory*, Academic Press.

Jacod, J., Y. Li, P. A. Mykland, M. Podolskij, and M. Vetter (2009): "Microstructure noise in the continuous case: The pre-averaging approach," *Stochastic Processes and Their Applications*, 119, 2249–2276.

Jacod, J., Y. Li, and X. Zheng (2019): "Estimating the integrated volatility with tick observations," *Journal of Econometrics*, 208, 80–100.

Jacod, J. and P. E. Protter (2011): *Discretization of Processes*, vol. 67, Springer Science & Business Media.

Jacod, J. and A. N. Shiryaev (2003): *Limit Theorems for Stochastic Processes*, vol. 288, Springer-Verlag Berlin.

Podolskij, M. and M. Vetter (2009a): "Bipower-type estimation in a noisy diffusion setting," *Stochastic Processes and Their Applications*, 119, 2803–2831.

——— (2009b): "Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps," *Bernoulli*, 15, 634–658.

Rio, E. (1997): "About the Lindeberg method for strongly mixing sequences," *ESAIM: Probability and Statistics*, 1, 35–61.